

# Query Answering over Contextualized RDF/OWL Knowledge with Forall-Existential Bridge Rules: Decidable Finite Extension Classes (Post Print)

Mathew Joseph<sup>1,2</sup>, Gabriel Kuper<sup>2</sup>, Till Mossakowski<sup>3</sup>, and Luciano Serafini<sup>1</sup>

<sup>1</sup> DKM, FBK-IRST, Trento, Italy

<sup>2</sup> DISI, University Of Trento, Trento, Italy

<sup>3</sup> Faculty of Computer Science, Otto-von-Guericke University of Magdeburg, 39016, Magdeburg, Germany

{joseph, kuper}@disi.unitn.it, {serafini}@fbk.eu,  
{mossakow}@iws.cs.uni-magdeburg.de

**Abstract.** The proliferation of contextualized knowledge in the Semantic Web (SW) has led to the popularity of knowledge formats such as *quads* in the SW community. A quad is an extension of an RDF triple with contextual information of the triple. In this paper, we study the problem of query answering over quads augmented with forall-existential bridge rules that enable interoperability of reasoning between triples in various contexts. We call a set of quads together with such expressive bridge rules, a quad-system. Query answering over quad-systems is undecidable, in general. We derive decidable classes of quad-systems, for which query answering can be done using forward chaining. Sound, complete and terminating procedures, which are adaptations of the well known chase algorithm, are provided for these classes for deciding query entailment. Safe, msafe, and csafe class of quad-systems restrict the structure of blank nodes generated during the chase computation process to be directed acyclic graphs (DAGs) of bounded depth. RR and restricted RR classes do not allow the generation of blank nodes during the chase computation process. Both data and combined complexity of query entailment has been established for the classes derived. We further show that quad-systems are equivalent to forall-existential rules whose predicates are restricted to ternary arity, modulo polynomial time translations. We subsequently show that the technique of safety, strictly subsumes in expressivity, some of the well known and expressive techniques, such as joint acyclicity and model faithful acyclicity, used for decidability guarantees in the realm of forall-existential rules.

**Keywords:** Contextualized Query Answering, Contextualized RDF/OWL knowledge bases, Multi-Context Systems, Quads, Query answering, forall-existential rules, Knowledge Representation, Semantic Web

## 1 Introduction

As the Semantic Web (SW) is getting more and more ubiquitous and its constellation of interlinked ontologies, the web of data, is seamlessly proliferating at a steady rate,

more and more applications have started using SW as a back end, providing their users manifold services, leveraging semantic technologies. One of the main reasons why SW enjoys such admirable hospitality from its mammoth geographically disparate users is its “simple” and “open” model. The model is simple, as the only intricacy that a creator/consumer of a SW application needs to be equipped with is that of a (RDF) triple. A triple  $t = (s, p, o)$  represents the most basic piece of knowledge in the SW, where  $s$ , called the *subject*, is an identifier for a person, place, thing, value, or a resource in general, about which the creator of  $t$  intended to express his/her knowledge using  $t$ .  $p$ , called the *predicate*, is an identifier for a property, attribute, or in general a binary relation that relates  $s$  with the component  $o$ , called the object, that is also an identifier for a resource similar to  $s$ . The model is called open, as it allows anybody, anywhere around the world to freely create their RDF/OWL ontologies about a domain of their choice, and publish them in (embedded) RDF/OWL formats in their web portals, also linking via URIs to the concepts in other similarly published ontologies. Thus the open model, in order to promote reuse and freedom, imposes no arbitration mechanism for the ontologies users publish on the SW.

A problem caused by this open model is that any piece of knowledge which a person publishes is often his/her own perspective about a particular domain, which largely is relative to this person. As a consequence, the truth value of a piece of knowledge in the SW is context-dependent. Recently, as a solution to the aforementioned problem, the SW community adopts the use of quads, an extension of triples, as the primary carrier of knowledge. A quad  $c: (s, p, o)$  thus adds a fourth component of the context  $c$  to the triple  $(s, p, o)$ , explicating the identifier of the context in which the triple holds. As a result, more and more triple-stores are becoming quad-stores. Some of the popular quad-stores are 4store<sup>1</sup>, Openlink Virtuoso<sup>2</sup>, and some of the currently popular triple-stores like Sesame<sup>3</sup>, Allegrograph<sup>4</sup> internally keep track of the contexts of triples. Some of the recent initiatives in this direction have also extended existing formats like N-Triples to N-Quads, which the RDF 1.1 has introduced as a W3C recommendation. The latest Billion triple challenge datasets have all been released in the N-Quads format.

Other benefits of quads over triples are that they allow knowledge creators to specify various attributes of meta-knowledge that further qualify knowledge [2], and also allow users to query for this meta knowledge [3]. These attributes, which explicate the various assumptions under which knowledge holds, are also called *context dimensions* [4]. Examples of context dimensions are provenance, creator, intended user, creation time, validity time, geo-location, and topic. Having defined knowledge that is contextualized, as in  $c_1: (\text{Renzi, primeMinsiterOf, Italy})$ , one can now declare in a meta-context  $mc$ , statements such as  $mc: (c_1, \text{creator, John})$ ,  $mc: (c_1, \text{expiryTime, "jun-2016"})$  that talk about the knowledge in context  $c_1$ , in this case its creator and expiry time. Another benefit of such a contextualized approach is that it opens possibilities of interesting ways for querying a contextualized knowledge base. For instance, if context  $c_1$  contains knowledge about football world cup 2014 and context  $c_2$  about football euro cup

<sup>1</sup> <http://4store.org>

<sup>2</sup> <http://virtuoso.openlinksw.com/rdf-quad-store/>

<sup>3</sup> <http://www.openrdf.org/>

<sup>4</sup> <http://www.franz.com/agraph/allegrograph/>

2012, then the query “who beat Italy in both world cup 2014 and euro cup 2012” can be formalized as the conjunctive query:

$$c_1: (x, \text{beat}, \text{Italy}) \wedge c_2: (x, \text{beat}, \text{Italy}),$$

where  $x$  is a variable.

When reasoning with knowledge in quad form, since knowledge can be grouped and divided context wise and simultaneously be fed to separate reasoning engines, this approach improves both efficiency and scalability [9]. Besides the above flexibility, *bridge rules* [5] can be provided for inter-operating the knowledge in different contexts. Such rules are primarily of the form:

$$c : \phi \rightarrow c' : \phi' \quad (1)$$

where  $\phi, \phi'$  are both atomic concept (role) symbols,  $c, c'$  are contexts. The semantics of such a rule is that if, for any  $\mathbf{a}$ ,  $\phi(\mathbf{a})$  holds in context  $c$ , then  $\phi'(\mathbf{a})$  should hold in context  $c'$ , where  $\mathbf{a}$  is a unary/binary vector depending on whether  $\phi, \phi'$  are concept/role symbols. Although such bridge rules serve the purpose of specifying knowledge interoperability from a source context  $c$  to a target context  $c'$ , in many practical situations there is the need of inter-operating multiple source contexts with multiple target contexts, for which the bridge rules of the form (1) are inadequate. Besides, one would also want the ability of creating new values in target contexts for the bridge rules.

In this work, we study contextual reasoning and query answering over contextualized RDF/OWL knowledge bases in the presence of *forall-existential bridge rules* that allow conjunctions and existential quantifiers in them, and hence are more expressive than those in DDL [5] and McCarthy et al. [6]. We provide a basic semantics for contextual reasoning based on which we provide procedures for conjunctive query answering. For query answering, we use the notion of a *distributed chase*, which is an extension of the standard *chase* [20, 21] that is widely used in the knowledge representation (KR) and Database (DB) settings for similar purposes. As far as the semantics for reasoning is concerned, we adopt the approach given in works such as Distributed Description Logics [5], E-connections [22], and two-dimensional logic of contexts [23], to use a set of interpretation structures as a model for contextualized knowledge. In this way, knowledge in each context is separately interpreted in a different interpretation structure. The main contributions of this work are:

1. We formulate a context-based semantics that reuses the standard RDF/OWL semantics, which can be used for reasoning over quad-systems. Studying conjunctive query answering over quad-systems, it turns out that the entailment problem of conjunctive queries is undecidable for the most general class of quad-systems, called *unrestricted quad-systems*.
2. We derive decidable subclasses of unrestricted quad-systems, namely *csafe*, *msafe*, and *safe* quad-systems, for which we detail both data and combined complexities of conjunctive query entailment. These classes are based on the constrained DAG structure of Skolem blank nodes generated during the chase construction. We also provide decision procedures to decide whether an input quad-system is safe (*csafe*, *msafe*) or not.

3. We further derive less expressive classes, RR and restricted RR quad-systems, for which no Skolem blank nodes are generated during the chase construction.
4. We show that the class of unrestricted quad-systems is equivalent to the class of ternary  $\forall\exists$  rule sets. We compare the derived classes of quad-systems with well known subclasses of  $\forall\exists$  rule sets, such as jointly acyclic and model faithful acyclic rule sets, and show that the technique of safety we propose, subsumes these other techniques, in expressivity.

The paper is structured as follows. In section 2, we formalize the idea of contextualized quad-systems, giving various definitions and notations for setting the background. In section 3, we formalize the problem of query answering for quad-systems, define notions such as distributed chase that are further used for query answering, and give the undecidability results of query entailment on unrestricted quad-systems. In section 4, we present *csafe*, *msafe*, and *safe* quad-systems and their computational properties. In section 5, RR quad-systems and restricted RR quad-systems are introduced. In section 6, we prove the equivalence of quad-systems with ternary  $\forall\exists$  rule sets, and formally compare a few well known decidable classes in the realm of  $\forall\exists$  rules to the classes of quad-systems, we presented in section 4. We provide a detailed discussion to other relevant related works in section 7, and conclude in section 8.

Note that parts of the contents of section 2 and section 3 has been taken from conference papers [11] and [12].

## 2 Contextualized Quad-Systems

In this section, we formalize the notion of a quad-system and its semantics. For any vector or sequence  $x$ , we denote by  $\|x\|$  the number of symbols in  $x$ , and by  $\{x\}$  the set of symbols in  $x$ . For any sets  $A$  and  $B$ ,  $A \rightarrow B$  denotes the set of all functions from set  $A$  to set  $B$ .

Given the set of URIs  $\mathbf{U}$ , the set of blank nodes  $\mathbf{B}$ , and the set of literals  $\mathbf{L}$ , the set  $\mathbf{C} = \mathbf{U} \uplus \mathbf{B} \uplus \mathbf{L}$  is called the set of (RDF) constants. Any  $(s, p, o) \in \mathbf{C} \times \mathbf{C} \times \mathbf{C}$  is called a generalized RDF triple (from now on, just triple). A graph is a set of triples. A *quad* is a tuple of the form  $c: (s, p, o)$ , where  $(s, p, o)$  is a triple and  $c$  is a URI<sup>5</sup>, called the *context identifier* that denotes the context of the RDF triple. A *quad-graph* is defined as a set of quads. For any quad-graph  $Q$  and any context identifier  $c$ , we denote by  $graph_Q(c)$  the set  $\{(s, p, o) | c: (s, p, o) \in Q\}$ . We denote by  $Q_{\mathcal{C}}$  the quad-graph whose set of context identifiers is  $\mathcal{C}$ . The set of constants occurring in  $Q_{\mathcal{C}}$  is given as  $\mathbf{C}(Q_{\mathcal{C}}) = \{c, s, p, o \mid c: (s, p, o) \in Q_{\mathcal{C}}\}$ . The set of URIs in  $Q_{\mathcal{C}}$  is given by  $\mathbf{U}(Q_{\mathcal{C}}) = \mathbf{C}(Q_{\mathcal{C}}) \cap \mathbf{U}$ . The set of blank nodes  $\mathbf{B}(Q_{\mathcal{C}})$  and the set of literals  $\mathbf{L}(Q_{\mathcal{C}})$  are similarly defined. Let  $\mathbf{V}$  be the set of variables, any element of the set  $\mathbf{C}^{\mathbf{V}} = \mathbf{V} \cup \mathbf{C}$  is a *term*. Any  $(s, p, o) \in \mathbf{C}^{\mathbf{V}} \times \mathbf{C}^{\mathbf{V}} \times \mathbf{C}^{\mathbf{V}}$  is called a *triple pattern*, and an expression of the form  $c: (s, p, o)$ , where  $(s, p, o)$  is a triple pattern,  $c$  a context identifier, is called a *quad pattern*. A triple pattern  $t$ , whose variables are elements of the vector  $x$  or elements of the vector  $y$  is written as  $t(x, y)$ . For any function  $f: A \rightarrow B$ , the *restriction* of  $f$  to

<sup>5</sup> Although, in general a context identifier can be a constant, for the ease of notation, we restrict them to be a URI

a set  $A'$ , is the mapping  $f|_{A'}$  from  $A' \cap A$  to  $B$  such that  $f|_{A'}(a) = f(a)$ , for each  $a \in A \cap A'$ . For any triple pattern  $t = (s, p, o)$  and function  $\mu$  from  $\mathbf{V}$  to a set  $A$ ,  $t[\mu]$  denotes  $(\mu'(s), \mu'(p), \mu'(o))$ , where  $\mu'$  is an extension of  $\mu$  to  $\mathbf{C}$  such that  $\mu'|_{\mathbf{C}}$  is the identity function. For any set of triple patterns  $G$ ,  $G[\mu]$  denotes  $\bigcup_{t \in G} t[\mu]$ . For any vector of constants  $\mathbf{a} = \langle a_1, \dots, a_{\|\mathbf{a}\|} \rangle$ , and vector of variables  $\mathbf{x}$  of the same length,  $\mathbf{x}/\mathbf{a}$  is the function  $\mu$  such that  $\mu(x_i) = a_i$ , for  $1 \leq i \leq \|\mathbf{a}\|$ . We use the notation  $t(\mathbf{a}, \mathbf{y})$  to denote  $t(\mathbf{x}, \mathbf{y})[\mathbf{x}/\mathbf{a}]$ . Similarly, the above notations are also extended to sets of quad-patterns. For instance  $Q(\mathbf{x}, \mathbf{y})$  denotes a set of quad-patterns, whose variables are from  $\mathbf{x}$  or  $\mathbf{y}$ , and  $Q(\mathbf{a}, \mathbf{y})$  is written for  $Q(\mathbf{x}, \mathbf{y})[\mathbf{x}/\mathbf{a}]$ . For the sake of interoperating knowledge in different contexts, bridge rules need to be provided:

*Bridge rules (BRs)* Formally, a BR is of the form:

$$\forall \mathbf{x} \forall \mathbf{z} [c_1: t_1(\mathbf{x}, \mathbf{z}) \wedge \dots \wedge c_n: t_n(\mathbf{x}, \mathbf{z}) \rightarrow \exists \mathbf{y} c'_1: t'_1(\mathbf{x}, \mathbf{y}) \wedge \dots \wedge c'_m: t'_m(\mathbf{x}, \mathbf{y})] \quad (2)$$

where  $c_1, \dots, c_n, c'_1, \dots, c'_m$  are context identifiers,  $\mathbf{x}, \mathbf{y}, \mathbf{z}$  are vectors of variables such that  $\{\mathbf{x}\}, \{\mathbf{y}\}$ , and  $\{\mathbf{z}\}$  are pairwise disjoint.  $t_1(\mathbf{x}, \mathbf{z}), \dots, t_n(\mathbf{x}, \mathbf{z})$  are triple patterns which do not contain blank-nodes, and whose set of variables are from  $\mathbf{x}$  or  $\mathbf{z}$ .  $t'_1(\mathbf{x}, \mathbf{y}), \dots, t'_m(\mathbf{x}, \mathbf{y})$  are triple patterns, whose set of variables are from  $\mathbf{x}$  or  $\mathbf{y}$ , and also does not contain blank-nodes. For any BR  $r$  of the form (2),  $body(r)$  is the set of quad patterns  $\{c_1: t_1(\mathbf{x}, \mathbf{z}), \dots, c_n: t_n(\mathbf{x}, \mathbf{z})\}$ , and  $head(r)$  is the set of quad patterns  $\{c'_1: t'_1(\mathbf{x}, \mathbf{y}), \dots, c'_m: t'_m(\mathbf{x}, \mathbf{y})\}$ , and the frontier of  $r$ ,  $fr(r) = \{\mathbf{x}\}$ . Occasionally, we also note the BR  $r$  above as  $body(r)(\mathbf{x}, \mathbf{z}) \rightarrow head(r)(\mathbf{x}, \mathbf{y})$ . The set of terms in a BR  $r$  is:

$$\mathbf{C}^V(r) = \{c, s, p, o \mid c: (s, p, o) \in body(r) \cup head(r)\}$$

The set of terms for a set of BRs  $R$  is  $\mathbf{C}^V(R) = \bigcup_{r \in R} \mathbf{C}^V(r)$ . The URIs, blank nodes, literals, variables of a BR  $r$  (resp. set of BRs  $R$ ) are similarly defined, and are denoted as  $\mathbf{U}(r), \mathbf{B}(r), \mathbf{L}(r), \mathbf{V}(r)$  (resp.  $\mathbf{U}(R), \mathbf{B}(R), \mathbf{L}(R), \mathbf{V}(R)$ ), respectively.

**Definition 1 (Quad-System).** A quad-system  $QS_C$  is defined as a pair  $\langle Q_C, R \rangle$ , where  $Q_C$  is a quad-graph, whose set of context identifiers is  $\mathcal{C}$ , and  $R$  is a set of BRs.

For any quad-system,  $QS_C = \langle Q_C, R \rangle$ , the set of constants in  $QS_C$  is given by  $\mathbf{C}(QS_C) = \mathbf{C}(Q_C) \cup \mathbf{C}(R)$ . The sets  $\mathbf{U}(QS_C), \mathbf{B}(QS_C), \mathbf{L}(QS_C)$ , and  $\mathbf{V}(QS_C)$  are similarly defined for any quad-system  $QS_C$ . For any quad-graph  $Q_C$  (BR  $r$ ), its symbol size  $\|Q_C\|$  ( $\|r\|$ ) is the number of symbols required to print  $Q_C$  ( $r$ ). Hence,  $\|Q_C\| \approx 4 * |Q_C|$ , where  $|Q_C|$  denotes the cardinality of the set  $Q_C$ . Note that  $|Q_C|$  equals the number of quads in  $Q_C$ . For a BR  $r$ ,  $\|r\| \approx 4 * k$ , where  $k$  is the number of quad-patterns in  $r$ . For a set of BRs  $R$ ,  $\|R\|$  is given as  $\sum_{r \in R} \|r\|$ . For any quad-system  $QS_C = \langle Q_C, R \rangle$ , its size  $\|QS_C\| = \|Q_C\| + \|R\|$ .

*Semantics* In order to provide a semantics for enabling reasoning over a quad-system, we need to use a local semantics for each context to interpret the knowledge pertaining to it. Since the primary goal of this paper is a decision procedure for query answering over quad-systems based on forward chaining, we consider the following desiderata for the choice of the local semantics and its deductive machinery:

- there exists an operation  $\text{lclosure}()$  that computes the deductive closure of a graph w.r.t to the local semantics using the local inference rules in a set LIR,
- each inference rule in LIR is *range restricted*, i.e. non value-generating,
- given a finite graph as input, the  $\text{lclosure}()$  operation terminates with a finite graph as output in polynomial time whose size is polynomial w.r.t. to the input set.

Some of the alternatives for the local semantics satisfying the above mentioned criterion are Simple, RDF, RDFS [31], OWL-Horst [27] etc. Assuming that a local semantics has been fixed, for any context  $c$ , we denote by  $I^c = \langle \Delta^c, \cdot^c \rangle$  an interpretation structure for the local semantics, where  $\Delta^c$  is the interpretation domain,  $\cdot^c$  the corresponding interpretation function. Also  $\models_{\text{local}}$  denotes the local satisfaction relation between a local interpretation structure and a graph. Given a quad graph  $Q_C$ , a *distributed interpretation structure* is an indexed set  $\mathcal{I}^C = \{I^c\}_{c \in \mathcal{C}}$ , where  $I^c$  is a local interpretation structure, for each  $c \in \mathcal{C}$ . We define the satisfaction relation  $\models$  between a distributed interpretation structure  $\mathcal{I}^C$  and a quad-system  $QS_C$  as:

**Definition 2 (Model of a Quad-System).** A distributed interpretation structure  $\mathcal{I}^C = \{I^c\}_{c \in \mathcal{C}}$  satisfies a quad-system  $QS_C = \langle Q_C, R \rangle$ , in symbols  $\mathcal{I}^C \models QS_C$ , iff all the following conditions are satisfied:

1.  $I^c \models_{\text{local}} \text{graph}_{Q_C}(c)$ , for each  $c \in \mathcal{C}$ ;
2.  $a^{c_i} = a^{c_j}$ , for any  $a \in \mathbf{C}$ ,  $c_i, c_j \in \mathcal{C}$ ;
3. for each BR  $r \in R$  of the form (2) and for each  $\sigma \in \mathbf{V} \rightarrow \Delta^c$ , where  $\Delta^c = \bigcup_{c \in \mathcal{C}} \Delta^c$ , if  $I^{c_1} \models_{\text{local}} t_1(\mathbf{x}, \mathbf{z})[\sigma], \dots, I^{c_n} \models_{\text{local}} t_n(\mathbf{x}, \mathbf{z})[\sigma]$ ,

then there exists a function  $\sigma' \supseteq \sigma$ , such that

$$I^{c'_1} \models_{\text{local}} t'_1(\mathbf{x}, \mathbf{y})[\sigma'], \dots, I^{c'_m} \models_{\text{local}} t'_m(\mathbf{x}, \mathbf{y})[\sigma'].$$

Condition 1 in the above definition ensures that for any model  $\mathcal{I}^C$  of a quad-graph, each  $I^c \in \mathcal{I}^C$  is a local model of the set of triples in context  $c$ . Condition 2 ensures that any constant  $c$  is rigid, i.e. represents the same resource across a quad-graph, irrespective of the context in which it occurs. Condition 3 ensures that any model of a quad-system satisfies each BR in it. Any  $\mathcal{I}^C$  such that  $\mathcal{I}^C \models QS_C$  is said to be a model of  $QS_C$ . A quad-system  $QS_C$  is said to be *consistent* if there exists a model  $\mathcal{I}^C$ , such that  $\mathcal{I}^C \models QS_C$ , and otherwise said to be *inconsistent*. For any quad-system  $QS_C = \langle Q_C, R \rangle$ , it can be the case that  $\text{graph}_{Q_C}(c)$  is locally consistent, i.e. there exists an  $I^c$  such that  $I^c \models_{\text{local}} \text{graph}_{Q_C}(c)$ , for each  $c \in \mathcal{C}$ , whereas  $QS_C$  is not consistent. This is because the set of BRs  $R$  adds more knowledge to the quad-system, and restricts the set of models that satisfy the quad-system.

**Definition 3 (Quad-system entailment).** (a) A quad-system  $QS_C$  entails a quad  $c: (s, p, o)$ , in symbols  $QS_C \models c: (s, p, o)$ , iff for any distributed interpretation structure  $\mathcal{I}^C$ , if  $\mathcal{I}^C \models QS_C$  then  $\mathcal{I}^C \models \langle \{c: (s, p, o)\}, \emptyset \rangle$ . (b) A quad-system  $QS_C$  entails a quad-graph  $Q'_{C'}$ , in symbols  $QS_C \models Q'_{C'}$ , iff  $QS_C \models c: (s, p, o)$  for any  $c: (s, p, o) \in Q'_{C'}$ . (c) A quad-system  $QS_C$  entails a BR  $r$  iff for any  $\mathcal{I}^C$ , if  $\mathcal{I}^C \models QS_C$  then  $\mathcal{I}^C \models \langle \emptyset, \{r\} \rangle$ . (d) For a set of BRs  $R$ ,  $QS_C \models R$  iff  $QS_C \models r$ , for every  $r \in R$ . (e) Finally, a quad-system  $QS_C$  entails another quad-system  $QS'_{C'} = \langle Q'_{C'}, R' \rangle$ , in symbols  $QS_C \models QS'_{C'}$ , iff  $QS_C \models Q'_{C'}$  and  $QS_C \models R'$ .

We call the decision problems corresponding to the entailment problems (EPs) in (a), (b), (c), (d), and (e) as *quad EP*, *quad-graph EP*, *BR EP*, *BRs EP*, and *quad-system EP*, respectively.

### 3 Query Answering on Quad-Systems

In the realm of quad-systems, the classical conjunctive queries or select-project-join queries are slightly extended to what we call *Contextualized Conjunctive Queries* (CCQs). A CCQ  $CQ(\mathbf{x})$  is an expression of the form:

$$\exists \mathbf{y} \, q_1(\mathbf{x}, \mathbf{y}) \wedge \dots \wedge q_p(\mathbf{x}, \mathbf{y}) \quad (3)$$

where  $q_i$ , for  $i = 1, \dots, p$  are quad patterns over vectors of *free variables*  $\mathbf{x}$  and *quantified variables*  $\mathbf{y}$ . A CCQ is called a boolean CCQ if it does not have any free variables. With some abuse, we sometimes discard the logical symbols in a CCQ and consider it as a set of quad-patterns. For any CCQ  $CQ(\mathbf{x})$  and a vector  $\mathbf{a}$  of constants such that  $\|\mathbf{x}\| = \|\mathbf{a}\|$ ,  $CQ(\mathbf{a})$  is boolean. A vector  $\mathbf{a}$  is an *answer* for a CCQ  $CQ(\mathbf{x})$  w.r.t. structure  $\mathcal{I}^C$ , in symbols  $\mathcal{I}^C \models CQ(\mathbf{a})$ , iff there exists assignment  $\mu: \{\mathbf{y}\} \rightarrow \mathbf{B}$  such that  $\mathcal{I}^C \models \bigcup_{i=1, \dots, p} q_i(\mathbf{a}, \mathbf{y})[\mu]$ . A vector  $\mathbf{a}$  is a *certain answer* for a CCQ  $CQ(\mathbf{x})$  over a quad-system  $QS_C$ , iff  $\mathcal{I}^C \models CQ(\mathbf{a})$ , for every model  $\mathcal{I}^C$  of  $QS_C$ . Given a quad-system  $QS_C$ , a CCQ  $CQ(\mathbf{x})$ , and a vector  $\mathbf{a}$ , decision problem of determining whether  $QS_C \models CQ(\mathbf{a})$  is called the *CCQ EP*. It can be noted that the other decision problems over quad-systems, namely Quad/Quad-graph EP, BR(s) EP, Quad-system EP, are reducible to the CCQ EP (See Property 6). Hence, in this paper, we primarily focus on the CCQ EP.

#### 3.1 dChase of a Quad-System

In order to build a procedure for query answering over a quad-system, we employ what has been called in the literature a *chase* [20, 21]. Specifically, we adopt notions of the *restricted chase* in Fagin et al. [24] (also called non-oblivious chase). In order to fit the framework of quad-systems, we extend the standard notion of chase to a *distributed chase*, abbreviated *dChase*. In the following, we show how the dChase of a quad-system can be constructed.

For a set of quad-patterns  $S$  and a set of terms  $T$ , we define the relation *T-connectedness* between quad-patterns in  $S$  as the least relation with:

- $q_1$  and  $q_2$  are *T-connected*, if  $\mathbf{C}^V(q_1) \cap \mathbf{C}^V(q_2) \cap T \neq \emptyset$ , for any two quad-patterns  $q_1, q_2 \in S$ ,
- if  $q_1$  and  $q_2$  are *T-connected*, and  $q_2$  and  $q_3$  are *T-connected*, then  $q_1$  and  $q_3$  are also *T-connected*, for any quad-patterns  $q_1, q_2, q_3 \in S$ .

It can be noted that *T-connectedness* is an equivalence relation and partitions  $S$  into a set of *T-components* (similar notion is called a *piece* in Baget et al. [15]). Note that for two distinct *T-components*  $P_1, P_2$  of  $S$ ,  $\mathbf{C}^V(P_1) \cap \mathbf{C}^V(P_2) \cap T = \emptyset$ . For any BR  $r = \text{body}(r)(\mathbf{x}, \mathbf{z}) \rightarrow \text{head}(r)(\mathbf{x}, \mathbf{y})$ , suppose  $P_1, P_2, \dots, P_k$  are the pairwise distinct  $\{\mathbf{y}\}$ -components of  $\text{head}(r)(\mathbf{x}, \mathbf{y})$ , then  $r$  can be replaced by the semantically equivalent set of BRs  $\{\text{body}(r)(\mathbf{x}, \mathbf{z}) \rightarrow P_1, \dots, \text{body}(r)(\mathbf{x}, \mathbf{z}) \rightarrow P_k\}$  whose symbol size is worst case quadratic w.r.t. the symbol size of  $r$ . Hence, w.l.o.g. we assume that for any BR  $r$ , the set of quad-patterns  $\text{head}(r)$  is a single component w.r.t. the set of existentially quantified variables in  $r$ .

Considering the fact that the local semantics for contexts are fixed a priori (for instance RDFS), both the number of rules in the set of local inference rules LIR and the size of each rule in LIR can be assumed to be a constant. Note that each local inference rule is range restricted and does not contain existentially quantified variables in its head. Any  $ir \in \text{LIR}$  is of the form:

$$\forall \mathbf{x} \forall \mathbf{z} [t_1(\mathbf{x}, \mathbf{z}) \wedge \dots \wedge t_k(\mathbf{x}, \mathbf{z}) \rightarrow t'_1(\mathbf{x})], \quad (4)$$

where  $t_i(\mathbf{x}, \mathbf{z})$ , for  $i = 1, \dots, n$  are triple patterns, whose variables are from  $\{\mathbf{x}\}$  or  $\{\mathbf{z}\}$ , and  $t'_1(\mathbf{x})$  is a triple pattern, whose variables are from  $\{\mathbf{x}\}$ . Hence, for any quad-system  $QS_C = \langle Q_C, R \rangle$  in order to accomplish the effect of local inferencing in each context  $c \in \mathcal{C}$ , for each  $ir \in \text{LIR}$  of the form (4), we could augment  $R$  with a BR  $ir_c$  of the form:

$$\forall \mathbf{x} \forall \mathbf{z} [c: t_1(\mathbf{x}, \mathbf{z}) \wedge \dots \wedge c: t_k(\mathbf{x}, \mathbf{z}) \rightarrow c: t'_1(\mathbf{x})]$$

Since  $\|\text{LIR}\|$  is a constant and the size of the augmentation is linear in  $|\mathcal{C}|$ , w.l.o.g we assume that the set  $R$  contains a BR  $ir_c$ , for each  $ir \in \text{LIR}$ ,  $c \in \mathcal{C}$ .

Given a quad-system  $QS_C$ , we denote by  $\mathbf{B}_{sk} \subseteq \mathbf{B}$ , a set of blank nodes called *Skolem blank nodes*, such that  $\mathbf{B}_{sk} \cap \mathbf{B}(QS_C) = \emptyset$ . For any BR  $r = \text{body}(r)(\mathbf{x}, \mathbf{z}) \rightarrow \text{head}(r)(\mathbf{x}, \mathbf{y})$  and an assignment  $\mu: \{\mathbf{x}\} \cup \{\mathbf{z}\} \rightarrow \mathbf{C}$ , the *application* of  $\mu$  on  $r$  is defined as:

$$\text{apply}(r, \mu) = \text{head}(r)[\mu^{ext}(\mathbf{y})]$$

where  $\mu^{ext}(\mathbf{y}) \supseteq \mu$  such that  $\mu^{ext}(\mathbf{y})(y_i) = \_$ :  $b$  is a fresh blank node from  $\mathbf{B}_{sk}$ , for each  $y_i \in \{\mathbf{y}\}$ .

We assume that there exists an order  $\prec_l$  (for instance, lexicographic order) on the set of constants. We extend  $\prec_l$  to the set of quads such that for any two quads  $c: (s, p, o)$  and  $c': (s', p', o')$ ,  $c: (s, p, o) \prec_l c': (s', p', o')$ , iff  $c \prec_l c'$ , or  $c = c', s \prec_l s'$ , or  $c = c', s = s', p \prec_l p'$ , or  $c = c', s = s', p = p', o \prec_l o'$ . It can be noted that  $\prec_l$  is a strict linear order over the set of all quads. For any finite quad-graph  $Q_C$ , the  $\prec_l$ -greatest quad of  $Q_C$ , denoted  $\text{greatestQuad}_{\prec_l}(Q_C)$ , is the quad  $q \in Q_C$  such that  $q' \prec_l q$ , for every other  $q' \in Q_C$ . Also, the order  $\prec_q$  is defined over the set of finite quad-graphs as follows: for any two finite quad-graphs  $Q_C, Q'_{C'}$ ,

$$\begin{aligned} Q_C \prec_q Q'_{C'}, & \text{ if (i) } Q_C \subset Q'_{C'}; \\ Q_C \prec_q Q'_{C'}, & \text{ if (i) does not hold and (ii) } \text{greatestQuad}_{\prec_l}(Q_C \setminus Q'_{C'}) \prec_l \text{greatestQuad}_{\prec_l}(Q'_{C'} \setminus Q_C); \\ Q_C \not\prec_q Q'_{C'}, & \text{ if both (i) and (ii) are not satisfied;} \end{aligned}$$

A relation  $R$  over a set  $A$  is called a *strict linear order* iff  $R$  is irreflexive, transitive, and  $R(a, b)$  or  $R(b, a)$  holds, for every distinct  $a, b \in A$ .

*Property 1.* Let  $\mathcal{Q}$  be the set of all finite quad-graphs;  $\prec_q$  is a strict linear order over  $\mathcal{Q}$ .

Also, we now define in parallel the dChase of a quad system  $QS_C = \langle Q_C, R \rangle$  and the level of a quad in the dChase of  $QS_C$  as follows: any quad in  $Q_C$  is of level 0. The level of a set of quads is the largest among levels of quads in the set. The level of any quad that results from the application of a BR  $r$  w.r.t. an assignment  $\mu$  is one more



than the level of the set  $body(r)[\mu]$ , if it has not already been assigned a level. Let  $\prec$  be an ordering on the quad-graphs such that for any two quad-graphs  $Q'_{C'}$  and  $Q''_{C''}$  of the same level,  $Q'_{C'} \prec Q''_{C''}$ , iff  $Q'_{C'} \prec_q Q''_{C''}$ . For  $Q'_{C'}$  and  $Q''_{C''}$  of different levels,  $Q'_{C'} \prec Q''_{C''}$ , iff level of  $Q'_{C'}$  is less than level of  $Q''_{C''}$ . It can easily be seen that  $\prec$  is a strict linear order over the set of quad-graphs. For any BRs  $r, r'$  and assignments  $\mu, \mu'$  over  $\mathbf{V}(body(r)), \mathbf{V}(body(r'))$ , respectively,  $(r, \mu) \prec (r', \mu')$  iff  $body(r)[\mu] \prec body(r')[\mu']$ . For any quad-graph  $Q'_{C'}$ , a set of BRs  $R$ , a BR  $r \in R$ , an assignment  $\mu \in \mathbf{V}(body(r)) \rightarrow \mathbf{C}$ , let  $applicable_R$  be the least ternary predicate defined inductively as:

$applicable_R(r, \mu, Q'_{C'})$  holds, if (a)  $body(r)[\mu] \subseteq Q'_{C'}$ ,  $head(r)[\mu'] \not\subseteq Q'_{C'}$ ,  $\forall \mu'' \supseteq \mu$ , and (b)  $\nexists r' \in R, \nexists \mu'$  such that  $r' \neq r$  or  $\mu' \neq \mu$  with  $(r', \mu') \prec (r, \mu)$  and  $applicable_R(r', \mu', Q'_{C'})$ ;

For any quad-system  $QS_C = \langle Q_C, R \rangle$ , let

$dChase_0(QS_C) = Q_C$ ;

$dChase_{i+1}(QS_C) = dChase_i(QS_C) \cup apply(r, \mu)$ , if there exists  $r = body(r)(x, z) \rightarrow head(r)(x, y) \in R$ , assignment  $\mu: \{x\} \cup \{z\} \rightarrow \mathbf{C}$  such that  $applicable_R(r, \mu, dChase_i(QS_C))$ ;

$dChase_{i+1}(QS_C) = dChase_i(QS_C)$ , otherwise; for any  $i \in \mathbb{N}$ . The dChase of  $QS_C$ , noted  $dChase(QS_C)$ , is given as:

$$dChase(QS_C) = \bigcup_{i \in \mathbb{N}} dChase_i(QS_C)$$

Intuitively,  $dChase_i(QS_C)$  can be thought of as the state of  $dChase(QS_C)$  at the end of iteration  $i$ . It can be noted that, if there exists  $i$  such that  $dChase_i(QS_C) = dChase_{i+1}(QS_C)$ , then  $dChase(QS_C)$  is equal to  $dChase_i(QS_C)$ . A model  $\mathcal{I}^C$  of a quad-system  $QS_C$  is called *universal* [30], iff the following holds:  $\mathcal{I}^C$  is a model of  $QS_C$ , and for any model  $\mathcal{I}'^C$  of  $QS_C$  there exists a homomorphism from  $\mathcal{I}^C$  to  $\mathcal{I}'^C$ .

**Theorem 1.** *For any consistent quad-system  $QS_C$ , the following holds: (i)  $dChase(QS_C)$  is a universal model of  $QS_C$ .<sup>6</sup>, and (ii) for any boolean  $CCQ \ CQ()$ ,  $QS_C \models CQ()$  iff there exists a map  $\mu: \mathbf{V}(CQ) \rightarrow \mathbf{C}$  such that  $\{CQ()\}[\mu] \subseteq dChase(QS_C)$ .*

An analog of the above theorem for DLs and Databases is stated and proved in [25]. Since the proof in [25] can easily be adapted to our case, we refer the reader to [25] for the proof. We call the sequence  $dChase_0(QS_C), dChase_1(QS_C), \dots$ , the *dChase sequence* of  $QS_C$ . It should be noted that at each iteration  $i$ , after the application of a BR, any new quad added is assigned a level, and as a result any subset of the set of quads in  $dChase_i(QS_C)$  has a level. This assignment of levels guarantees that  $applicable_R(r, \mu, dChase_i(QS_C))$  is either true or false, for any  $r \in R$ , assignment  $\mu: \mathbf{V}(body(r)) \rightarrow \mathbf{C}$ . The following lemma shows that in a dChase sequence of a quad-system, any dChase iteration can be performed in time exponential w.r.t. the size of the largest BR.

<sup>6</sup> Though  $dChase(QS_C)$  is not an interpretation in a strict model theoretic sense, one can easily create the corresponding interpretation  $\mathcal{I}_{dChase(QS_C)} = \{I^c = \langle \Delta^c, \cdot^c \rangle\}_{c \in \mathcal{C}}$ , s.t. for every  $c \in \mathcal{C}$ ,  $\Delta^c$  is equal to set of constants in  $graph_{dChase(QS_C)}(c)$ , and  $\cdot^c$  is s.t  $(s, p, o) \in graph_{dChase(QS_C)}(c)$  iff  $(s^c, o^c) \in p^c$ .

**Lemma 1.** For a quad-system  $QS_C = \langle Q_C, R \rangle$ , for any  $i \in \mathbb{N}^+$ , the following holds:  
(i)  $dChase_i(QS_C)$  can be computed in time  $\mathcal{O}(|R| * \|dChase_{i-1}(QS_C)\|^{rs})$ , where  $rs = \max_{r \in R} \|r\|$ , (ii)  $\|dChase_i(QS_C)\| = \mathcal{O}(\|dChase_{i-1}(QS_C)\| + \|R\|)$ .

*Proof.* (i) We can first find, if there exists an  $r$  among the set of BRs  $R$ , assignment  $\mu$  such that  $applicable_R(r, \mu, dChase_{i-1}(QS_C))$  holds, in the following naive way: (1) bind the set of variables in all rules in  $R$  with the set of constants in  $dChase_{i-1}(QS_C)$ . Let this set be called  $S$ . Note that  $|S| = \mathcal{O}(|R| * \|dChase_{i-1}(QS_C)\|^{rs})$ , where  $rs = \max_{r \in R} \|r\|$ . Also, note that each of the binding in  $S$  is of the form  $body(r)(\mathbf{x}, \mathbf{z})(\mu) \rightarrow head(r)(\mathbf{x}, \mathbf{y})(\mu') (\heartsuit)$ , where  $r \in R$ . (2) From the set  $S$  we filter out every binding of the form  $(\heartsuit)$  in which  $\mathbf{x}[\mu] \neq \mathbf{x}[\mu']$ . Let  $S'$  be the resulting set after the above filtering operation. (3) From the set  $S'$ , we now filter out all the bindings of the form  $(\heartsuit)$  with  $head(r)(\mathbf{x}, \mathbf{y})(\mu') \subseteq dChase_{i-1}(QS_C)$ , with resulting set  $S''$ . (4) If  $S'' = \emptyset$ , then there is no  $r \in R$ , assignment  $\mu$  such that  $applicable_R(r, \mu, dChase_{i-1}(QS_C))$  is True. Otherwise if  $S'' \neq \emptyset$ , then note that each binding of the form  $(\heartsuit)$  in  $S''$  is such that condition (a) of the true  $applicable_R(r, \mu, dChase_{i-1}(QS_C))$  is satisfied. Now, we can sort  $S''$  w.r.t.  $\prec$  and select the least binding  $b$  of the form  $(\heartsuit)$ , so that condition (b) in True condition of  $applicable_R()$  is satisfied for  $b$ . It can easily be seen that  $applicable_R(r, \mu, dChase_{i-1}(QS_C))$  holds for the  $r, \mu$  extracted from  $b$ . Since the size of each binding is at most  $\|rs\|$ , the operations (1)-(4) can be performed in time  $\mathcal{O}(|R| * \|dChase_{i-1}(QS_C)\|^{rs})$ . Since  $dChase_i(QS_C) = dChase_{i-1}(QS_C) \cup head(r)[\mu]$ , for  $r, \mu$  with  $applicable_R(r, \mu, dChase_{i-1}(QS_C))$ ,  $dChase_i(QS_C)$  can be computed in time  $\mathcal{O}(\|dChase_{i-1}(QS_C)\|^{rs})$ .

(ii) Trivially holds, since at worst  $dChase_i(QS_C) = dChase_{i-1}(QS_C) \cup head(r)[\mu]$ , for  $r \in R$ .

**Lemma 2.** For any quad-system  $QS_C$ , If  $\_:$   $b$  is a Skolem blank node in  $dChase(QS_C)$ , generated by the application of assignment  $\mu$  on  $r = body(r)(\mathbf{x}, \mathbf{z}) \rightarrow head(r)(\mathbf{x}, \mathbf{y})$ , with  $\mu^{ext(\mathbf{y})}(y_j) = \_:$   $b, y_j \in \{\mathbf{y}\}$ , then  $\_:$   $b$  is unique for  $(r, y_j, \mathbf{x}[\mu^{ext(\mathbf{y})}])$ .

*Proof.* By contradiction, suppose if  $\_:$   $b$  is not unique for  $(r, y_j, \mathbf{x}[\mu^{ext(\mathbf{y})}])$ , i.e. there exists  $\_:$   $b' \neq \_:$   $b$  in  $dChase(QS_C)$ , with  $\_:$   $b'$  generated by  $r$  such that  $\_:$   $b' = \mu'^{ext(\mathbf{y})}(y_j)$  and  $\mathbf{x}[\mu^{ext(\mathbf{y})}] = \mathbf{x}[\mu'^{ext(\mathbf{y})}]$ . W.l.o.g. suppose  $\_:$   $b$  was generated in an iteration  $l \in \mathbb{N}$  and  $\_:$   $b'$  in an iteration  $m > l$ . This means that  $head(r)(\mathbf{x}, \mathbf{y})[\mu^{ext(\mathbf{y})}] \subseteq dChase_l(QS_C)$ , and hence  $head(r)(\mathbf{x}, \mathbf{y})[\mu^{ext(\mathbf{y})}] \subseteq dChase_{m-1}(QS_C)$ . Also, since  $\mu|_{\mathbf{x}} = \mu'|_{\mathbf{x}}$ , there  $\exists \mu'' \supseteq \mu'$  s.t.  $head(r)(\mathbf{x}, \mathbf{y})[\mu''] \subseteq dChase_{m-1}(QS_C)$ . This means that (a) part of the function  $applicable_R$  is false, for  $applicable_R(r, \mu', dChase_{m-1}(QS_C))$  to be true, and hence  $applicable_R(r, \mu', dChase_{m-1}(QS_C))$  is false. Hence, our assumption that  $\_:$   $b' = y_j[\mu'^{ext(\mathbf{y})}]$  is false. Hence,  $\_:$   $b$  is unique for  $(r, y_j, \mathbf{x}[\mu^{ext(\mathbf{y})}])$ .

Although we now know how to compute the dChase of a quad-system, which can be used for deciding CCQ EP, the following proposition reveals that for the class of quad-systems whose BRs are of the form (2), which we call *unrestricted quad-systems*, the dChase can be infinite.

**Proposition 1.** There exists unrestricted quad-systems whose dChase is infinite.

*Proof.* Consider an example of a quad-system  $QS_c = \langle Q_c, r \rangle$ , where  $Q_c = \{c: (a, \text{rdf:type}, C)\}$ , and the BR  $r = c: (x, \text{rdf:type}, C) \rightarrow \exists y c: (x, P, y), c: (y, \text{rdf:type}, C)$ . The dChase computation starts with  $dChase_0(QS_c) = \{c: (a, \text{rdf:type}, C)\}$ , now the rule  $r$  is applicable, and its application leads to  $dChase_1(QS_c) = \{c: (a, \text{rdf:type}, C), c: (a, P, \_ : b_1), c: (\_ : b_1, \text{rdf:type}, C)\}$ , where  $\_ : b_1$  is a fresh Skolem blank node. It can be noted that  $r$  is yet again applicable on  $dChase_1(QS_c)$ , for  $c: (\_ : b_1, \text{rdf:type}, C)$ , which leads to the generation of another Skolem blank node, and so on. Hence,  $dChase(QS_c)$  does not have a finite fix-point, and  $dChase(QS_c)$  is infinite.

A class  $\mathcal{C}$  of quad-systems is called a *finite extension class* (FEC), iff for every member  $QS_c \in \mathcal{C}$ ,  $dChase(QS_c)$  is a finite set. Therefore, the class of unrestricted quad-systems is not a FEC. This raises the question if there are other approaches that can be used, for instance, a similar problem of non-finite chase is manifested in description logics (DLs) with value creation, due to the presence of existential quantifiers, whereas the approaches like the one in Glimm et al. [28] provides an algorithm for CQ entailment based on query rewriting. Theorem 2 below establishes the fact that the CCQ EP for unrestricted quad-systems is undecidable. Despite this, the reader should note that the following undecidability result and its proof is only provided for the sake of self containedness, and we do not claim the undecidability theorem nor its proof to be a novel contribution, as we will show in section 6, ternary  $\forall\exists$  rule sets are polynomially reducible to unrestricted quad-systems. Hence, the undecidability results provided in Baget et al. [15], Krötzsch et al. [39], or Beeri et al. [14] can trivially be applied in our setting to obtain the undecidability result for unrestricted quad-systems.

**Theorem 2.** *The CCQ entailment problem over unrestricted quad-systems is undecidable.*

*Proof.* (sketch) We show that the well known undecidable problem of non-emptiness of intersection of context-free grammars (CFGs) is reducible to the CCQ entailment problem. Given two CFGs,  $G_1 = \langle V_1, T, S_1, P_1 \rangle$  and  $G_2 = \langle V_2, T, S_2, P_2 \rangle$ , where  $V_1, V_2$  are the set of variables,  $T$  such that  $T \cap (V_1 \cup V_2) = \emptyset$  is the set of terminals.  $S_1 \in V_1$  is the start symbol of  $G_1$ , and  $P_1$  are the set of PRs of the form  $v \rightarrow w$ , where  $v \in V$ ,  $w$  is a sequence of the form  $w_1 \dots w_n$ , where  $w_i \in V_1 \cup T$ .  $S_2, P_2$  are defined similarly. Deciding whether the language generated by the grammars  $L(G_1)$  and  $L(G_2)$  have non-empty intersection is known to be undecidable [33].

Given two CFGs  $G_1 = \langle V_1, T, S_1, P_1 \rangle$  and  $G_2 = \langle V_2, T, S_2, P_2 \rangle$ , we encode grammars  $G_1, G_2$  into a quad-system  $QS_c = \langle Q_c, R \rangle$ , with only a single context identifier  $c$ . Each PR  $r = v \rightarrow w \in P_1 \cup P_2$ , with  $w = w_1 w_2 w_3 \dots w_n$ , is encoded as a BR of the form:  $c: (x_1, w_1, x_2), c: (x_2, w_2, x_3), \dots, c: (x_n, w_n, x_{n+1}) \rightarrow c: (x_1, v, x_{n+1})$ , where  $x_1, \dots, x_{n+1}$  are variables. For each terminal symbol  $t_i \in T$ ,  $R$  contains a BR of the form:  $c: (x, \text{rdf:type}, C) \rightarrow \exists y c: (x, t_i, y), c: (y, \text{rdf:type}, C)$  and  $Q_c$  is the singleton:  $\{c: (a, \text{rdf:type}, C)\}$ . It can be observed that:

$$QS_c \models \exists y c: (a, S_1, y) \wedge c: (a, S_2, y) \Leftrightarrow L(G_1) \cap L(G_2) \neq \emptyset$$

We refer the reader to Appendix for the complete proof.

Having shown the undecidability results of query answering of unrestricted quad-systems, the rest of the paper focuses on defining subclasses of unrestricted quad-systems for which query answering is decidable, and establishing their relationships with similar classes in the realm of  $\forall\exists$  rules. While defining decidable classes for quad-systems, one mainly has two fundamentally distinct options: (i) is to define notions that solely use the structure/properties of the BR part, ignoring the quad-graph part, or (ii) to define notions that takes into account both the BR and quad-graph part. The decidability notions which we define in section 4, namely safety, msafety, and csafety belong to type (ii), as these techniques take into account the property of the dChase of a quad-system, which is determined by both the quad-graph and BRs of the quad-system. Whereas the ones which we define in section 5, namely RR and restricted RR quad-systems fall into type (i), as the properties of BRs alone are used. With an analogy between a set of BRs and a set of  $\forall\exists$  rules, and between a quad-graph and a set of  $\forall\exists$  instances, the reader should note that such distinctions can also be made for the decidability notions in the realm of  $\forall\exists$  rule sets. Techniques such as Weak acyclicity [24], Joint acyclicity [38], and Acyclic graph of rule dependencies [15] belong to type (ii), as these notions ignore the instance part. Whereas techniques such as model faithful acyclicity [32] and model summarizing acyclicity [32] are of type (i) as both the rules and instance part is considered.

#### 4 Safe, Msafe and Csafe Quad-Systems: Decidable FECs

In the previous section, we saw that the query answering problem over unrestricted quad-systems is undecidable, in general. We will also see in section 6 that any quad-system is polynomially translatable to a  $\forall\exists$  rule set, which is also a first order logic theory. Hence, a possible solution approach is to translate to these more expressive languages, and apply well known tests (see related work for details on such tests) available in these languages to check if query answering is decidable. If the translated quad-system passes one of these tests, then query answering can be performed on this translation using available algorithms in these expressive languages. But such an approach is often discouraged, because of the non-applicability of the already available tools and techniques available for reasoning over quads. Instead, we in the following define three classes of quad-systems, namely SAFE, MSAFE and CSAFE, that are FECs and for which query entailment is decidable. Finiteness/decidability is achieved by putting certain restrictions (explained below) on the blank nodes generated in the dChase.

Recall that, for any quad-system  $QS_C$ , the set of blank-nodes  $\mathbf{B}(dChase(QS_C))$  in its  $dChase(QS_C)$  not only contains blank nodes present in  $QS_C$ , i.e.  $\mathbf{B}(QS_C)$ , but also contains Skolem blank nodes that are generated during the dChase construction process. Note that the following holds:  $\mathbf{B}_{sk}(dChase(QS_C)) = \mathbf{B}(dChase(QS_C)) \setminus \mathbf{B}(QS_C)$ . We assume w.l.o.g. that for any set of BRs  $R$ , any BR in  $R$  has a unique rule identifier, and we often write  $r_i$  for the BR in  $R$ , whose identifier is  $i$ .

**Definition 4 (Origin RuleId/Vector).** *For any Skolem blank node  $\_ : b$ , generated in the dChase by the application of a BR  $r_i = \text{body}(r_i)(\mathbf{x}, \mathbf{z}) \rightarrow \text{head}(r_i)(\mathbf{x}, \mathbf{y})$  using assignment  $\mu : \{\mathbf{x}\} \cup \{\mathbf{z}\} \rightarrow \mathbf{C}$ , i.e.  $b = \mu^{ext(\mathbf{y})}(y_j)$ , for some  $y_j \in \mathbf{y}$ , we say that the origin ruleId of  $\_ : b$  is  $i$ , denoted  $\text{originRuleId}(\_ : b) = i$ . Moreover  $\mathbf{w} = \mathbf{x}[\mu]$  is said to be the origin vector of  $\_ : b$ , denoted  $\text{originVector}(\_ : b) = \mathbf{w}$ .*

As we saw in Lemma 2, any such Skolem blank node  $\_ : b$ , generated in the dChase can uniquely be represented by the expression  $(i, j, w)$ , where  $i$  is rule id,  $j$  is identifier of the existentially quantified variable  $y_j$  in  $r_i$  substituted by  $\_ : b$  during the application of  $\mu$  on  $r_i$ . Also in the above case, we denote relation between each constant  $k = \mu^{ext(y)}(x_h)$ ,  $x_h \in \{x\}$ , and  $\_ : b$  with the relation *childOf*. Moreover, since children of a Skolem blank node can be Skolem blank nodes, which themselves can have children, one can naturally define relation *descendantOf* = *childOf*<sup>+</sup> as the transitive closure of *childOf*. Note that according to the above definition, ‘descendantOf’ is not reflexive. In addition, we could keep track of the set of contexts in which a blank-node was first generated, using the following notion:

**Definition 5 (Origin-contexts).** For any quad-system  $QS_C$  and for any Skolem blank node  $\_ : b \in \mathbf{B}_{sk}(dChase(QS_C))$ , the set of origin-contexts of  $\_ : b$  is given by *origin-Contexts*( $\_ : b$ ) =  $\{c \mid \exists i. c:(s, p, o) \in dChase_i(QS_C), s = \_ : b \text{ or } p = \_ : b \text{ or } o = \_ : b, \text{ and } \nexists j < i \text{ with } c':(s', p', o') \in dChase_j(QS_C), s' = \_ : b \text{ or } p' = \_ : b \text{ or } o' = \_ : b, \text{ for any } c' \in \mathcal{C}\}$ .

Intuitively, origin-contexts for a Skolem blank node  $\_ : b$  is the set of contexts in which triples containing  $\_ : b$  are first generated, during the dChase construction. Note that there can be multiple contexts in which  $\_ : b$  can simultaneously be generated. By setting *originRuleId*( $k$ ) = *n.d.*, (resp. *originVector*( $k$ ) = *n.d.*, resp. *originContexts*( $k$ ) = *n.d.*) where *n.d.* is an ad hoc constant,  $\forall k \notin \mathbf{B}_{sk}(dChase(QS_C))$ , we extend the definition of origin ruleId, (resp. origin vector, resp. origin-contexts) to all the constants in the dChase of a quad-system.

*Example 1.* Consider the quad-system  $\langle Q_C, R \rangle$ , where  $Q_C = \{c_1 : (a, b, c)\}$ . Suppose  $R$  is the following set:

$$R = \left\{ \begin{array}{ll} c_1 : (x_{11}, x_{12}, z_1) \rightarrow c_2 : (x_{11}, x_{12}, y_1) & (r_1) \\ c_2 : (a, z_2, x_{22}) \rightarrow c_3 : (a, x_{22}, y_2) & (r_2) \\ c_2 : (z_3, b, x_{32}) \rightarrow c_3 : (b, x_{32}, y_3) & (r_3) \\ c_3 : (a, z_{41}, x_{41}), c_3 : (b, z_{42}, x_{42}) & \\ \rightarrow c_2 : (y_4, x_{41}, a), c_2 : (y_4, x_{42}, b) & (r_4) \end{array} \right\}$$

Suppose that for brevity quantifiers have been omitted, and variables of the form  $y_i$  or  $y_{ij}$  are implicitly existentially quantified. Iterations during the dChase construction are:

$$\begin{aligned} dChase_0(QS_C) &= \{c_1 : (a, b, c)\} \\ dChase_1(QS_C) &= \{c_1 : (a, b, c), c_2 : (a, b, \_ : b_1)\} \\ dChase_2(QS_C) &= \{c_1 : (a, b, c), c_2 : (a, b, \_ : b_1), c_3 : (a, \_ : b_1, \_ : b_2)\} \\ dChase_3(QS_C) &= \{c_1 : (a, b, c), c_2 : (a, b, \_ : b_1), c_3 : (a, \_ : b_1, \_ : b_2), \\ &\quad c_3 : (b, \_ : b_1, \_ : b_3)\} \\ dChase_4(QS_C) &= \{c_1 : (a, b, c), c_2 : (a, b, \_ : b_1), c_3 : (a, \_ : b_1, \_ : b_2), c_3 : (b, \_ : b_1, \\ &\quad \_ : b_3), c_2 : (\_ : b_4, \_ : b_2, a), c_2 : (\_ : b_4, \_ : b_3, b)\} \\ dChase_5(QS_C) &= dChase_4(QS_C), \end{aligned}$$

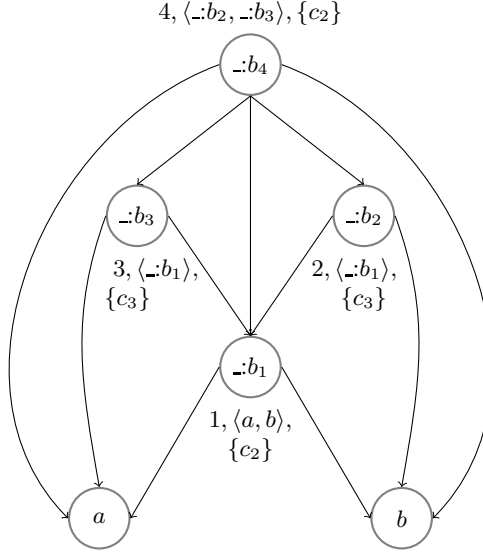


Fig. 1: descentance graph of  $\_ : b_4$  in example 1. Note: n.d. labels not shown

Also note:

$originRuleId(\_ : b_1) = 1$ ,  $originRuleId(\_ : b_2) = 2$ ,  $originRuleId(\_ : b_3) = 3$ ,  
 $originRuleId(\_ : b_4) = 4$ ,  
 $originVector(\_ : b_1) = \langle a, b \rangle$ ,  $originVector(\_ : b_2) = originVector(\_ : b_3) = \langle \_ : b_1 \rangle$ ,  
 $originVector(\_ : b_4) = \langle \_ : b_2, \_ : b_3 \rangle$ ,  
 also  $originContexts(\_ : b_1) = \{c_2\}$ ,  $originContexts(\_ : b_2) = originContexts(\_ : b_3) = \{c_3\}$ ,  
 $originContexts(\_ : b_4) = \{c_2\}$ ,  
 also  $\_ : b_1$  descendantOf  $\_ : b_3$ ,  $\_ : b_1$  descendantOf  $\_ : b_2$ ,  $\_ : b_2$  descendantOf  $\_ : b_4$ ,  $\_ : b_3$   
 descendantOf  $\_ : b_4$ ,  $\_ : b_1$  descendantOf  $\_ : b_4$ .

For any Skolem blank node  $\_ : b$  (in dChase), its descendant hierarchy can be analyzed using a *descentance graph*  $\langle V, E, \lambda_r, \lambda_v, \lambda_c \rangle$ , which is a labeled graph rooted at  $\_ : b$ , whose set of nodes  $V$  are constants in the dChase, the set of edges  $E$  is such that  $(k, k') \in E$ , iff  $k'$  is a descendant of  $k$ .  $\lambda_r, \lambda_v, \lambda_c$  are node labeling functions, such that  $\lambda_r(k) = originRuleId(k)$ ,  $\lambda_v(k) = originVector(k)$ , and  $\lambda_c(k) = originContexts(k)$ , for any  $k \in V$ . The descentance graph for  $\_ : b_4$  of Example 1 is shown in Fig.1. For any two vectors of constants  $v, w$ , we note  $v \cong w$ , iff there exists a bijection  $\mu : \mathbf{B}(v) \rightarrow \mathbf{B}(w)$  such that  $w = v[\mu]$ .

**Definition 6 (safe, msafe, csafe quad-systems).** A quad-system  $QS_C$  is said to be unsafe (resp. unmsafe, resp. unsafe), iff there exist Skolem blank nodes  $\_ : b \neq \_ : b'$  in  $dChase(QS_C)$  such that  $\_ : b$  is a descendant of  $\_ : b'$ , with  $originRuleId(\_ : b) = originRuleId(\_ : b')$  and  $originVector(\_ : b) \cong originVector(\_ : b')$  (resp.  $originRuleId(\_ : b) = originRuleId(\_ : b')$ , resp.  $originContexts(\_ : b) = originContexts(\_ : b')$ ).

$xts(\_ : b')$ ). A quad-system is safe (resp. msafe, resp. csafe) iff it is not unsafe (resp. unmsafe, resp. unsafe).

Intuitively, safe, msafe and csafe quad-systems, does not allow repetitive generation of Skolem blank-nodes with a certain set of attributes in its dChase. The containment relation between the class of safe, msafe, and csafe quad-systems are established by the following theorem:

**Theorem 3.** Let SAFE, MSAFE, and CSAFE denote the class of safe, msafe, and csafe quad-systems, respectively, then the following holds:

$$CSAFE \subset MSAFE \subset SAFE$$

*Proof.* We first show  $MSAFE \subseteq SAFE$ , by showing the inverse inclusion of their compliments, i.e.  $UNSAFE \subseteq UNMSAFE$ . Suppose a given quad-system  $QS_C$  is unsafe, then by definition its dChase contains two distinct Skolem blank nodes  $\_ : b, \_ : b'$  such that  $\_ : b$  is a descendant of  $\_ : b'$ , with  $originRuleId(\_ : b) = originRuleId(\_ : b')$  and  $originVector(\_ : b) \cong originVector(\_ : b')$ . But this will imply that  $originRuleId(\_ : b) = originRuleId(\_ : b')$ . Hence, by definition,  $QS_C$  is unmsafe. Hence  $UNSAFE \subseteq UNMSAFE$  ( $\dagger$ ).

Now, we show that  $CSAFE \subseteq MSAFE$  by showing  $UNMSAFE \subseteq UNCSAFE$ . Suppose a given quad-system  $QS_C = \langle Q_C, R \rangle$  is unmsafe, then by definition its dChase contains two distinct Skolem blank nodes  $\_ : b, \_ : b'$  such that  $\_ : b$  is a descendant of  $\_ : b'$ , with  $originRuleId(\_ : b) = originRuleId(\_ : b')$ . But this implies that there exists a BR  $r_i = body(r_i)(\mathbf{x}, \mathbf{z}) \rightarrow head(r_i)(\mathbf{x}, \mathbf{y})$ , assignment  $\mu$ , (resp.  $\mu'$ ), s.t.  $\_ : b$  (resp.  $\_ : b'$ ) was generated in  $dChase(QS_C)$  as result of application of  $\mu$  (resp.  $\mu'$ ) on  $r_i$ . That is  $\_ : b = y_j[\mu^{ext}(\mathbf{y})]$ , and  $\_ : b' = y_k[\mu'^{ext}(\mathbf{y})]$ , where  $y_j, y_k \in \{\mathbf{y}\}$ . We have the following two subcases (i)  $j = k$ , (ii)  $j \neq k$ . Suppose (i)  $j = k$ , then it immediately follows that  $originContexts(\_ : b) = originContexts(\_ : b')$ . Hence,  $QS_C$  is unsafe. Suppose (ii)  $j \neq k$ , then by construction of dChase, on application of  $\mu'$  to  $r_i$ , along with  $\_ : b'$ , there gets also generated a Skolem blank node  $\_ : b'' = y_j[\mu'^{ext}(\mathbf{y})]$ , with  $y_j \in \{\mathbf{y}\}$ . Since  $\_ : b$  and  $\_ : b''$  are generated by substitutions of the same variable  $y_j \in \{\mathbf{y}\}$  of BR  $r_i$ ,  $originContexts(\_ : b) = originContexts(\_ : b'')$ . Also considering that  $childOf(\_ : b') = childOf(\_ : b'') = \{\mathbf{x}[\mu^{ext}(\mathbf{y})]\}$ , we can deduce that  $\_ : b$  is a descendant of  $\_ : b''$ . Hence, by definition, it holds that  $QS_C$  is unsafe. Hence  $UNMSAFE \subseteq UNCSAFE$  ( $\ddagger$ ).

From  $\dagger$  and  $\ddagger$ , it follows that  $CSAFE \subseteq MSAFE \subseteq SAFE$ . To show that the containments are strict, consider the quad-system  $QS_C$  in example 1. By definition,  $QS_C$  is msafe, however unsafe, as the Skolem blank nodes  $\_ : b_1, \_ : b_4$ , which have the same origin contexts are s.t.  $\_ : b_1$  is a descendant of  $\_ : b_4$ . Hence,  $CSAFE \subset MSAFE$ . For  $MSAFE \subset SAFE$ , the following example shows an instance of a quad-system that is unmsafe, yet is safe.

**Example 2.** Consider the quad-system  $QS_C = \langle Q_C, R \rangle$ , where  $Q_C = \{c_1 : (a, b, c), c_2 : (c, d, e)\}$ ,  $R$  is given by:

$$c_1 : (x_{11}, x_{12}, x_{13}), c_2 : (x_{13}, x_{14}, z_1) \rightarrow c_3 : (y_1, x_{11}, x_{12}), c_4 : (x_{12}, x_{13}, x_{14}) \quad (r_1)$$

$$c_3 : (x_{21}, a, x_{22}), c_4 : (x_{22}, x_{23}, x_{24}) \rightarrow c_1 : (x_{21}, a, x_{22}), c_2 : (x_{22}, x_{23}, x_{24}) \quad (r_2)$$

$$c_3 : (x_{21}, x_{22}, a), c_4 : (a, x_{23}, x_{24}) \rightarrow c_1 : (x_{21}, x_{22}, a), c_2 : (a, x_{23}, x_{24}) \quad (r_3)$$

$$c_3 : (x_{21}, x_{22}, x_{23}), c_4 : (x_{23}, a, x_{24}) \rightarrow c_1 : (x_{21}, x_{22}, x_{23}), c_2 : (x_{23}, a, x_{24}) \quad (r_4)$$

$$c_3: (x_{21}, x_{22}, x_{23}), c_4: (x_{23}, x_{24}, a) \rightarrow c_1: (x_{21}, x_{22}, x_{23}), c_2: (x_{23}, x_{24}, a) \quad (r_5)$$

Note that for brevity quantifiers have been omitted, and variables of the form  $y_i$  or  $y_{ij}$  are implicitly existentially quantified. Iterations during the dChase construction are:

$$\begin{aligned} dChase_0(QS_C) &= \{c_1: (a, b, c), c_2: (c, d, e)\} \\ dChase_1(QS_C) &= dChase_0(QS_C) \cup \{c_3: (\_ : b_1, a, b), c_4: (b, c, d)\} \\ dChase_2(QS_C) &= dChase_1(QS_C) \cup \{c_1: (\_ : b_1, a, b), c_2: (b, c, d)\} \\ dChase_3(QS_C) &= dChase_2(QS_C) \cup \{c_3: (\_ : b_2, \_ : b_1, a), c_4: (a, b, c)\} \\ dChase_4(QS_C) &= dChase_3(QS_C) \cup \{c_1: (\_ : b_2, \_ : b_1, a), c_2: (a, b, c)\} \\ dChase_5(QS_C) &= dChase_4(QS_C) \cup \{c_3: (\_ : b_3, \_ : b_2, \_ : b_1), c_4: (\_ : b_1, a, b)\} \\ dChase_6(QS_C) &= dChase_5(QS_C) \cup \{c_1: (\_ : b_3, \_ : b_2, \_ : b_1), c_2: (\_ : b_1, a, b)\} \\ dChase_7(QS_C) &= dChase_6(QS_C) \cup \{c_3: (\_ : b_4, \_ : b_3, \_ : b_2), c_4: (\_ : b_2, \_ : b_1, a)\} \\ dChase_8(QS_C) &= dChase_7(QS_C) \cup \{c_1: (\_ : b_4, \_ : b_3, \_ : b_2), c_2: (\_ : b_2, \_ : b_1, a)\} \\ dChase_9(QS_C) &= dChase_8(QS_C) \cup \{c_3: (\_ : b_5, \_ : b_4, \_ : b_3), c_4: (\_ : b_3, \_ : b_2, \_ : b_1)\} \\ dChase(QS_C) &= dChase_9(QS_C) \end{aligned}$$

It can be seen that  $\_ : b_1, \_ : b_2, \_ : b_3, \_ : b_4, \_ : b_5$  form a descendant chain, since  $\_ : b_i$  descendantOf  $\_ : b_{i+1}$ , for each  $i = 1, \dots, 4$ . Also,  $originRuleId(\_ : b_i) = originRuleId(\_ : b_{i+1})$ , for each  $i = 1, \dots, 4$ . Hence it turns out that  $QS_C$  is unmsafe. However, it can be seen that  $originVector(\_ : b_1) = \langle a, b, c, d \rangle$ , and  $originVector(\_ : b_2) = \langle \_ : b_1, a, b, c \rangle$ , and  $originVector(\_ : b_3) = \langle \_ : b_2, \_ : b_1, a, b \rangle$ , and  $originVector(\_ : b_4) = \langle \_ : b_3, \_ : b_2, \_ : b_1, a \rangle$ , and  $originVector(\_ : b_5) = \langle \_ : b_4, \_ : b_3, \_ : b_2, \_ : b_1 \rangle$ , and  $originVector(\_ : b_i) \not\cong originVector(\_ : b_j)$ , for  $1 \leq i \neq j \leq 5$ , and hence, by definition,  $QS_C$  is safe with a terminating dChase. It can be noticed that during each distinct application of  $r_1$ , the vector of constants bound to the vector of variables  $\langle x_{11}, \dots, x_{14} \rangle$  are different w.r.t  $\cong$ . Safe quad-systems in this way are capable of recognizing such positive cases of finite dChases (which are classified as negative cases by msafe quad-systems) by also keeping track of the origin vectors of Skolem blank-nodes in their dChases.

The following property shows that for a safe quad-system, the descendant graph of any Skolem blank node in its dChase is a directed acyclic graph (DAG):

*Property 2 (DAG property).* For a safe (csafe, msafe) quad-system  $QS_C$ , and for any blank node  $b \in \mathbf{B}_{sk}(dChase(QS_C))$ , its descendant graph is a DAG.

*Proof.* By construction, as there exists no descendant for any constant  $k \in \mathbf{C}(QS_C)$ , there cannot be any out-going edge from any such  $k$ . Hence, no member of  $\mathbf{C}(QS_C)$  can be involved in cycles. Therefore, the only members that can be involved in cycles are the members of  $\mathbf{C}(dChase(QS_C)) - \mathbf{C}(QS_C) = \mathbf{B}_{sk}(dChase(QS_C))$ . But if there exists  $\_ : b \in \mathbf{B}_{sk}(dChase(QS_C))$ , such that there exists a cycle through  $\_ : b$ , then this implies that  $\_ : b$  is a descendant of  $\_ : b$ . This would violate the prerequisites of being safe (resp. csafe, resp. msafe), and imply that  $QS_C$  is unsafe (resp. unsafe, resp. unmsafe), which is a contradiction.



---

**Algorithm 1:**

---

```
UnRavel (Descendance Graph  $G$ )
/* procedure to unravel, a descendance graph into a tree */
Input : descendance graph  $G = \langle V, E, \lambda_r, \lambda_v, \lambda_c \rangle$ 
Output: A labeled Tree  $G$ 
begin
   $G = \langle V, E, \lambda_r, \lambda_v, \lambda_c \rangle := \text{RemoveTranstiveEdges}(G)$ ;
  foreach Node  $v_o \in \text{preOrder}(G)$  do
    if ( $k = \text{indegree}(v_o) > 1$ ) then
       $\{v_1, \dots, v_k\} := \text{getFreshNodes}()$ ; /* each  $v_i \notin V$  is fresh */
      /* replace old node  $v_o$  by the fresh nodes in  $V$  */
      removeNodeFrom( $v_o, V$ );
      addNodesTo( $\{v_1, \dots, v_k\}, V$ );
      foreach  $(v_o, v') \in E$  do
        /* replace each outgoing edge from  $v_o$  with a fresh outgoing
           edges from each fresh node  $v_i$  */
        removeEdgeFrom( $(v_o, v'), E$ );
        addEdgesTo( $\{(v_1, v'), \dots, (v_k, v')\}, E$ );
       $i := 1$ ;
      foreach  $(v', v_o) \in E$  do
        /* replace each incoming edge of  $v_o$  with an incoming edge for
           a unique  $v_i$  */
        removeEdgeFrom( $(v', v_o), E$ );
        addEdgeTo( $(v', v_i), E$ );
         $i++$ ;
  /* restrict node labels to the updated set of nodes in  $V$  */
   $\lambda_r := \lambda_r|_V, \lambda_v := \lambda_v|_V, \lambda_c := \lambda_c|_V$ ;
  return  $G$ ;
```

---

Since the descendance graph  $G$  of any Skolem blank node  $\_$ :  $b \in \mathbf{B}_{sk}(dChase(QS_C))$  is such that  $G$  is rooted at  $\_$ :  $b$  and is acyclic, any directed path from  $\_$ :  $b$  terminates at some node. Hence, one can use a tree traversal technique, such as preorder (visit a node first and then its children) to sequentially traverse nodes in  $G$ . Algorithm 1 takes a descendance graph  $G$  and unravels it into a tree. The algorithm first removes all the transitive edges from  $G$ , i.e. if there are  $v, v' \in V$  with  $(v, v') \in E$  and  $G$  contains a path of length greater than 1 from  $v$  to  $v'$ , then it removes  $(v, v')$ . Note that, in the resulting graph, the presence of a path from  $v$  to  $v''$  still gives us the information that  $v''$  is a descendant of  $v$ . The algorithm then traverses the graph in preorder fashion, as it encounters a node  $v$ , if  $v$  has an indegree  $k$  greater than one, it replaces  $v$  with  $k$  fresh nodes  $v_1, \dots, v_k$ , and distributes the set of edges incident to  $v$  across  $v_1, \dots, v_k$ , such that (i) each  $v_i$  has at-most one incoming edge (ii) all the edges incident to  $v$  are incident to some  $v_i$ ,  $i \in \{1, \dots, k\}$ . Outgoing edges of  $v$  are copied for each  $v_i$ . Hence, after the above operation each  $v_i$  has an indegree 1, whereas outdegree of  $v_i$  is same as the outdegree of  $v$ ,  $i \in \{1, \dots, k\}$ . Hence, after all the nodes are visited, every node except the root in the new graph  $G$  has an indegree 1.  $G$  is still rooted, connected, acyclic, and is hence a tree. The algorithm terminates as there are no cycles in the graph, and at some point reaches a node with no children. For instance, the unraveling of the descendance graph of  $\_$ :  $b_4$  in Fig. 1 is shown in Fig. 2. The following property holds for any Skolem blank node of a safe quad-system.

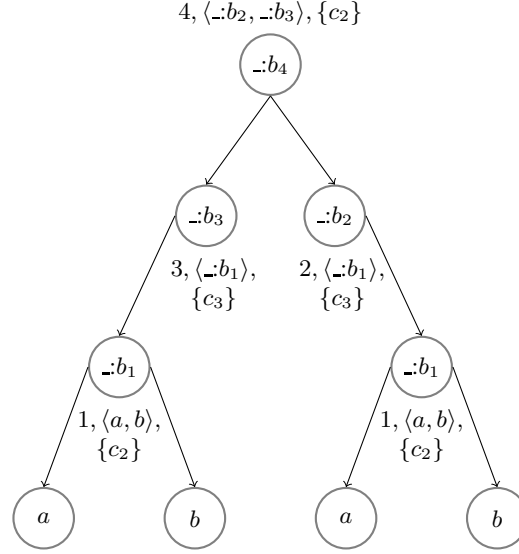


Fig. 2: Descentance graph of Fig. 1 unraveled into a tree. Note: n.d. labels are not shown

*Property 3.* For a safe quad-system  $QS_C = \langle Q_C, R \rangle$ , and any Skolem blank node in  $dChase(QS_C)$ , the unraveling (Algorithm 1) of its descentance graph results in a tree  $t = \langle V, E, \lambda_r, \lambda_v, \lambda_c \rangle$  s.t.:

1. any leaf node of  $t$  is from the set  $\mathbf{C}(QS_C)$ ,
2. any non-leaf node of  $t$  is from  $\mathbf{B}_{sk}(dChase(QS_C))$ ,
3.  $order(t) \leq w$ , where  $w = \max_{r \in R} |fr(r)|$ ,
4. there cannot be a path between  $b \neq b' \in V$ , with  $\lambda_r(b) = \lambda_r(b')$  and  $\lambda_v(b) \cong \lambda_v(b')$ ,
5. there cannot be a path between  $b \neq b' \in V$ , with  $\lambda_r(b) = \lambda_r(b')$ , if  $QS_C$  is also msafe,
6. there cannot be a path between  $b \neq b' \in V$ , with  $\lambda_c(b) = \lambda_c(b')$ , if  $QS_C$  is also csafe.

*Proof.* 1. Any node  $n$  in the descentance graph is such that  $n \in \mathbf{C}(dChase(QS_C))$ , and  $\mathbf{C}(dChase(QS_C)) = \mathbf{C}(QS_C) \uplus \mathbf{B}_{sk}(dChase(QS_C))$ . Since any member  $m \in \mathbf{B}_{sk}(dChase(QS_C))$  is generated from an application of a BR with an assignment  $\mu$  such that its frontier variables are assigned by  $\mu$  with a set of constants,  $m$  has at-least one child. But, since  $n$  is a leaf node,  $n \in \mathbf{C}(QS_C)$ .

2. Since no member  $m \in \mathbf{C}(QS_C)$  can have descendants and any non-leaf node has children,  $m$  cannot be a non-leaf node. Hence, non-leaf nodes must be from  $\mathbf{B}_{sk}(dChase(QS_C))$ .

3. The order of  $t$  is the maximal outdegree among the nodes of  $t$ , and outdegree of a node is the number of children it has. Since any node in  $t$  with non-zero outdegree is a Skolem blank-node  $\therefore b$  generated by application of an assignment  $\mu$  to  $r =$

$body(r)(x, z) \rightarrow head(r)(x, y) \in R$ , the number of children  $z$  of  $b$  has equals  $\|x\|$ . Hence the order of  $t$  is bounded by  $w$ .

4. Since any path from  $b$  to  $b'$  implies that  $b'$  is a descendant of  $b$ , it must be the case that  $\lambda_r(b) \neq \lambda_r(b')$  or  $\lambda_v(b) \not\cong \lambda_v(b')$ , otherwise safety condition would be violated.
5. Similar as above, immediate by definition.
6. Similar as above, immediate by definition.

The property above is exploited to show that there exists a finite bound in the dChase size and its computation time.

**Lemma 3.** *For any safe/msafe/csafes quad-system  $QS_C = \langle Q_C, R \rangle$ , the following holds: (i) the dChase size  $\|dChase(QS_C)\| = \mathcal{O}(2^{2^{\|QS_C\|}})$ , (ii)  $dChase(QS_C)$  can be computed in 2EXPTIME, (iii) if  $\|R\|$  and the set of schema triples in  $Q_C$  is fixed to a constant, then  $\|dChase(QS_C)\|$  is a polynomial in  $\|QS_C\|$  and can be computed in PTIME.*

*Proof.* The proofs are provided for safe quad-systems, but since CSAFES  $\subset$  MSAFES  $\subset$  SAFES and since we are giving upper bounds, they also propagate trivially to msafe and csafes quad-systems.

(i) For any Skolem blank node in  $dChase(QS_C)$ , the size of its originVector is upper bounded by  $w = \max_{r \in R} |fr(r)|$ . If  $S$  is the set of all origin vectors of blank-nodes in  $dChase(QS_C)$ , then cardinality of the set  $S' = S \setminus \cong$  is upper bounded by  $(|U(QS_C)| + |L(QS_C)| + w)^w$ , which means that  $|S'| = \mathcal{O}(2^{\|QS_C\|})$ . Also, since the set of origin ruleId labels,  $Rids$ , can at most be  $|R|$ , the cardinality of the set  $Rids \times S' = \mathcal{O}(2^{\|QS_C\|})$ . For the descendence tree  $t$  of any Skolem blank node of  $dChase(QS_C)$ , since there cannot be paths in  $t$  between distinct  $b$  and  $b'$ , such that  $originRuleId(b) = originRuleId(b')$  and  $originVector(b) \cong originVector(b')$ , the length of any such path is upper bounded by  $|Rids \times S'| = \mathcal{O}(2^{\|QS_C\|})$ . However, it turns out that the above upper bound provided is loose, as there is the need of additional *filter* BRs to transform/back-propagate vectors of constants associated with Skolem blank nodes generated by repetitive application of the same BR. For instance, consider the set of BRs in eg: 2. The BR  $r_1$  transforms the origin vector to a new vector each time during its application. BRs  $r_2 - r_5$  deals with back propagation of these vectors back to input origin vectors of BR  $r_1$ . Such filter BRs rule out the case of a BR being applied to a quad that contains a Skolem blank node that was generated using the same BR on an isomorphic origin vector, ensuring that the safety criteria for Skolem blank-nodes generated is not violated. It turns out that the number of such filter BRs required is polynomial w.r.t. to the number of descendants with the same rule id, for a node in  $t$ . Hence, it turns out the depth of  $t$  is polynomially bounded by  $\|R\|$ . (Note that depth of  $t$  is bounded by  $|R|$  for msafe quad-systems. Also since, the set of origin context labels are bounded by the set of existential variables in  $R$ , depth of  $t$  is bounded by  $\|R\|$  for csafes quad-systems.) Also order of the tree is bounded by  $w$ . Hence, any such tree can have at most  $\mathcal{O}(2^{\|QS_C\|})$  leaf nodes,  $\mathcal{O}(2^{\|QS_C\|})$  inner nodes, and  $\mathcal{O}(2^{\|QS_C\|})$  nodes. Since each of the leaf nodes can only be from  $C(QS_C)$  and each of the inner nodes correspond to an existential variable in  $R$ , the number of such possible trees are clearly

bounded double exponentially in  $\|QS_C\|$ , hence bounds the number of Skolem blank nodes generated in the dChase.

(ii) From (i)  $\|dChase(QS_C)\|$  is double exponential in  $\|QS_C\|$ , and since each iteration add at-least one quad to its dChase, the number of iterations are bounded double exponentially in  $\|QS_C\|$ . Also, by Lemma 1 any iteration  $i$  can be done in time  $\mathcal{O}(\|dChase_{i-1}(QS_C)\|^{\|R\|})$ . Hence, by using (i), we get  $\|dChase_{i-1}(QS_C)\| = \mathcal{O}(2^{2^{\|QS_C\|}})$ . Hence, we can infer that each iteration  $i$  can be done in time  $\mathcal{O}(2^{\|R\| * 2^{\|QS_C\|}})$ . Also since the number of iterations is at most double exponential, computing  $dChase(QS_C)$  is in 2EXPTIME.

(iii) Since  $\|R\|$  is fixed to a constant, the set of existential variables is also a constant. In this case, since the size of the frontier of any  $r \in R$  is also a constant, the order and depth of any descendant tree  $t$  of a Skolem blank node is a constant. Hence, the number of (leaf) nodes of  $t$  is bounded by a constant. Also in this setting, the label of inner nodes of  $t$ , which correspond to existential variables, is also a constant, and the leaf nodes of  $t$  can only be a constant in  $\mathbf{C}(QS_C)$ . Hence, the number of descendant trees and consequently, the number of Skolem blank nodes generated is bounded by  $\mathcal{O}(|\mathbf{C}(QS_C)|^z)$ , where  $z$  is a constant. Hence, the set of constants generated in  $dChase(QS_C)$  is a polynomial in  $\|QS_C\|$ , and so is  $\|dChase(QS_C)\|$ .

Since in any dChase iteration except the final one, at least one quad is added, and also since the final dChase can have at most  $\mathcal{O}(\|QS_C\|^z)$  triples, the total number of iterations are bounded by  $\mathcal{O}(\|QS_C\|^z)$  (†). By Lemma 1, since any iteration  $i$  can be computed in  $\mathcal{O}(\|dChase_{i-1}(QS_C)\|^{\|R\|})$  time, and since  $\|R\|$  is a constant, the time required for each iteration is a polynomial in  $\|dChase_{i-1}(QS_C)\|$ , which is at most a polynomial in  $\|QS_C\|$ . Hence, any dChase iteration can be performed in polynomial time in size of  $QS_C$  (‡). From (†) and (‡), it can be concluded that dChase can be computed in PTIME.

**Lemma 4.** *For any safe/msafe/csafes quad-system, the following holds: (i) data complexity of CCQ entailment is in PTIME, (ii) combined complexity of CCQ entailment is in 2EXPTIME.*

*Proof.* Note that the proofs are provided for safe quad-systems, but since CSAFES  $\subset$  MSAFES  $\subset$  SAFES and since we are giving upper bounds, they also propagate trivially to msafe and csafes quad-systems.

Given a safe quad-system  $QS_C = \langle Q_C, R \rangle$ , since  $dChase(QS_C)$  is finite, a boolean CCQ  $CQ()$  can naively be evaluated by binding the set of constants in the dChase to the variables in the  $CQ()$ , and then checking if any of these bindings are contained in  $dChase(QS_C)$ . The number of such bindings can at most be  $\|dChase(QS_C)\|^{\|CQ()\|}$  (†).

(i) Since for data complexity, the size of the BRs  $\|R\|$ , the set of schema triples, and  $\|CQ()\|$  is fixed to a constant. From Lemma 3 (iii), we know that under the above mentioned settings the dChase can be computed in PTIME and is polynomial in the size of  $QS_C$ . Since  $\|CQ()\|$  is fixed to a constant, and from (†), binding the set of constants in  $dChase(QS_C)$  on  $CQ()$  still gives a number of bindings that is worst case polynomial in the size of  $\|QS_C\|$ . Since membership of these bindings can be checked in the polynomially sized dChase in PTIME, the time required for CCQ entailment is in PTIME.

(ii) Since in this case  $\|dChase(QS_C)\| = \mathcal{O}(2^{2^{\|QS_C\|}})$  ( $\dagger$ ), from ( $\dagger$ ) and ( $\ddagger$ ), binding the set of constants in  $dChase(QS_C)$  to  $CQ()$  amounts to  $\mathcal{O}(2^{\|CQ()\| * 2^{\|QS_C\|}})$  number of bindings. Since the dChase is double exponential in  $\|QS_C\|$ , checking the membership of each of these bindings can be done in 2EXPTIME. Hence, the combined complexity is in 2EXPTIME.

**Theorem 4.** *For any safe/msafe/csaf quad-system, the following holds: (i) The data complexity of CCQ entailment is PTIME-complete (ii) The combined complexity of CCQ entailment is 2EXPTIME-complete.*

*Proof.* (i)(Membership) See Lemma 4 for the membership in PTIME.

(Hardness) Follows from the PTIME-hardness of data complexity of CCQ entailment for Range-Restricted quad-systems (Theorem 8), which are contained in safe/msafe/csaf quad-systems.

(ii) (Membership) See Lemma 4.

(Hardness) See following heading.

#### 4.1 2EXPTIME-Hardness of CCQ Entailment

In this subsection, we show that the combined complexity of the decision problem of CCQ entailment for context acyclic quad-systems is 2EXPTIME-hard. We show this by reduction of the word-problem of a 2EXPTIME deterministic turing machine (DTM) to the CCQ entailment problem. We notify the reader that the technique we follow is, similar to works such as [35, 36], to iteratively generate a doubly exponential number of objects that represent the configurations and cells of the tape of the DTM, and then simulate its working by appropriate BRs. A DTM  $M$  is a tuple  $M = \langle Q, \Sigma, \Delta, q_0, q_A \rangle$ , where

- $Q$  is a set of states,
- $\Sigma$  is a finite set of letters that includes the blank symbol  $\square$ ,
- $\Delta: (Q \times \Sigma) \rightarrow (Q \times \Sigma \times \{+1, -1\})$  is the transition function,
- $q_0 \in Q$  is the initial state.
- $q_A \in Q$  is the accepting state.

W.l.o.g. we assume that there exists exactly one accepting state, which is also the lone halting state. A configuration is a word  $\alpha \in \Sigma^* Q \Sigma^*$ . A configuration  $\alpha_2$  is a successor of the configuration  $\alpha_1$ , iff one of the following holds:

1.  $\alpha_1 = w_l q \sigma \sigma_r w_r$  and  $\alpha_2 = w_l \sigma' q' \sigma_r w_r$ , if  $\Delta(q, \sigma) = (q', \sigma', R)$ , or
2.  $\alpha_1 = w_l q \sigma$  and  $\alpha_2 = w_l \sigma' q' \square$ , if  $\Delta(q, \sigma) = (q', \sigma', R)$ , or
3.  $\alpha_1 = w_l \sigma_l q \sigma w_r$  and  $\alpha_2 = w_l q' \sigma_l \sigma' w_r$ , if  $\Delta(q, \sigma) = (q', \sigma', L)$ .

where  $q, q' \in Q$ ,  $\sigma, \sigma', \sigma_l, \sigma_r \in \Sigma$ , and  $w_l, w_r \in \Sigma^*$ . Since number of configurations can at most be doubly exponential in the size of the input string, and since  $2EXPTIME \subseteq 2EXPSPACE$ , the number of tape cells traversed by the DTM tape head is also bounded double exponentially. A configuration  $c = w_l q w_r$  is an accepting configuration iff  $q = q_A$ . A language  $L \subseteq \Sigma^*$  is accepted by a 2EXPTIME bounded DTM  $M$ , iff for every  $w \in L$ ,  $M$  accepts  $w$  in time  $\mathcal{O}(2^{2^{\|w\|}})$ .

*Simulating DTMs using Safe Quad-Systems* Consider a double exponential time bounded DTM  $M = \langle Q, \Sigma, \Delta, q_0, q_A \rangle$ , and a string  $w$ , with  $\|w\| = m$ . Suppose that  $M$  terminates in  $2^{2^n}$  time, where  $n = m^k$ ,  $k$  is a constant. In order to simulate  $M$ , we construct a quad-system  $QS_C^M = \langle Q_C^M, R \rangle$ , where  $C = \{c_0, c_1, \dots, c_n\}$ , whose various elements represents the constructs of  $M$ . Let  $Q_C^M$  be initialized with the following quads:

$$c_0 : (k_0, \text{rdf} : \text{type}, R), c_0 : (k_1, \text{rdf} : \text{type}, R), c_0 : (k_0, \text{rdf} : \text{type}, \text{min}_0), \\ c_0 : (k_1, \text{rdf} : \text{type}, \text{max}_0), c_0 : (k_0, \text{succ}_0, k_1)$$

Now for each pair of elements of type  $R$  in  $c_i$ , a Skolem blank-node is generated in  $c_{i+1}$ , and hence follows the recurrence relation  $r(j+1) = [r(j)]^2$ , with seed  $r(0) = 2$ , which after  $n$  iterations yields  $2^{2^n}$ . In this way, a doubly exponentially long chain of elements is created in  $c_n$  using the following set of rules:

$$c_i : (x_0, \text{rdf} : \text{type}, R), c_i : (x_1, \text{rdf} : \text{type}, R) \rightarrow \exists y \ c_{i+1} : (x_0, x_1, y), \\ c_{i+1} : (y, \text{rdf} : \text{type}, R) \quad (eBr)$$

The combination of the minimal element with the minimal element (elements of type  $\text{min}_i$ ) in  $c_i$  create the minimal element in  $c_{i+1}$ , and similarly the combination of the maximal element with the maximal element (elements of type  $\text{max}_i$ ) in  $c_i$  create the maximal element of  $c_{i+1}$ :

$$c_{i+1} : (x_0, x_0, x_1), c_i : (x_0, \text{rdf} : \text{type}, \text{min}_i) \rightarrow c_{i+1} : (x_1, \text{rdf} : \text{type}, \text{min}_{i+1}) \\ c_{i+1} : (x_0, x_0, x_1), c_i : (x_0, \text{rdf} : \text{type}, \text{max}_i) \rightarrow c_{i+1} : (x_1, \text{rdf} : \text{type}, \text{max}_{i+1})$$

The successor relation  $\text{succ}_{i+1}$  is created in  $c_{i+1}$  using the following set of rules, using the well-known integer counting technique:

$$c_i : (x_1, \text{succ}_i, x_2), c_{i+1} : (x_0, x_1, x_3), c_{i+1} : (x_0, x_2, x_4) \rightarrow c_{i+1} : (x_3, \text{succ}_{i+1}, x_4) \\ c_i : (x_1, \text{succ}_i, x_2), c_{i+1} : (x_1, x_3, x_5), c_{i+1} : (x_2, x_4, x_6), c_i : (x_3, \text{rdf} : \text{type}, \text{max}_i), \\ c_i : (x_4, \text{rdf} : \text{type}, \text{min}_i) \rightarrow c_{i+1} : (x_5, \text{succ}_{i+1}, x_6)$$

Each of the above set of rules are instantiated for  $0 \leq i < n$ , and in this way after  $n$  generating dChase iterations,  $c_n$  has doubly exponential number of elements of type  $R$ , that are ordered linearly using the relation  $\text{succ}_n$ . By virtue of the first rule below, each of the objects representing the cells of the DTM are linearly ordered by the relation  $\text{succ}$ . Also the transitive closure of  $\text{succ}$  is defined as the relation  $\text{succt}$

$$c_n : (x_0, \text{succ}_n, x_1) \rightarrow c_n : (x_0, \text{succ}, x_1) \\ c_n : (x_0, \text{succ}, x_1) \rightarrow c_n : (x_0, \text{succt}, x_1) \\ c_n : (x_0, \text{succt}, x_1), c_n : (x_1, \text{succt}, x_2) \rightarrow c_n : (x_0, \text{succt}, x_2)$$

Also using a similar construction, we can reuse the  $2^{2^{n-1}}$  linearly ordered elements in  $c_{n-1}$  to create another linearly ordered chain of a doubly exponential number of objects in  $c_n$  that represents configurations of  $M$ , whose minimal element is of type  $\text{conInit}$ , and the linear order relation being  $\text{conSucc}$ .

Various triple patterns that are used to encode the possible configurations, runs and their relations in  $M$  are:

$(x_0, head, x_1)$  denotes the fact that in configuration  $x_0$ , the head of the DTM is at cell  $x_1$ .  
 $(x_0, state, x_1)$  denotes the fact that in configuration  $x_0$ , the DTM is in state  $x_1$ .  
 $(x_0, \sigma, x_1)$  where  $\sigma \in \Sigma$ , denotes the fact that in configuration  $x_0$ , the cell  $x_1$  contains  $\sigma$ .  
 $(x_0, succ, x_1)$  denotes the linear order between cells of the tape.  
 $(x_0, succt, x_1)$  denotes the transitive closure of  $succ$ .  
 $(x_0, conSucc, x_1)$  to denote the fact that  $x_1$  is a successor configuration of  $x_0$ .  
 $(x_0, rdf : type, Accept)$  denotes the fact that the configuration  $x_0$  is an accepting configuration.

Since in our construction, each  $\sigma \in \Sigma$  is represented as a relation, we could constrain that no two letters  $\sigma \neq \sigma'$  are on the same cell using the following axiom:

$$c_n : (z_1, \sigma, z_2), c_n : (z_1, \sigma', z_2) \rightarrow$$

for each  $\sigma \neq \sigma' \in \Sigma$ . Note that the above BR has an empty head, is equivalent to asserting the negation of its body.

*Initialization* Suppose the initial configuration is  $q_0 w \square$ , where  $w = \sigma_0 \dots \sigma_{n-1}$ , then we enforce this using the following BRs in our quad-system  $QS_C^M$  as:

$$\begin{aligned}
 & c_n : (x_0, rdf : type, conInit), c_n : (x_1, rdf : type, min_n) \rightarrow c_n : (x_0, head, x_1), \\
 & c_n : (x_0, state, q_0) \\
 & c_n : (x_0, rdf : type, min_n) \wedge \bigwedge_{i=0}^{n-1} c_n : (x_i, succ, x_{i+1}) \wedge c_n : (x_j, rdf : type, \\
 & conInit) \rightarrow \bigwedge_{i=0}^{n-1} c_n : (x_j, \sigma_i, x_i) \wedge c_n : (x_j, \square, x_n) \\
 & c_n : (x_j, rdf : type, conInit), c_n : (x_j, \square, x_0), c_n : (x_0, succt, x_1) \rightarrow c_n : (x_j, \square, x_1)
 \end{aligned}$$

The last BR copies the  $\square$  to every succeeding cell in the initial configuration.

*Transitions* For every left transition  $\Delta(q, \sigma) = (q_j, \sigma', -1)$ , the following BR:

$$\begin{aligned}
 & c_n : (x_0, head, x_i), c_n : (x_0, \sigma, x_i), c_n : (x_0, state, q), c_n : (x_j, succ, x_i), c_n : (x_0, \\
 & conSucc, x_1) \rightarrow c_n : (x_1, head, x_j), c_n : (x_1, \sigma', x_i), c_n : (x_1, state, q_j)
 \end{aligned}$$

For every right transition  $\Delta(q, \sigma) = (q_j, \sigma', +1)$ , the following BR:

$$\begin{aligned}
 & c_n : (x_0, head, x_i), c_n : (x_0, \sigma, x_i), c_n : (x_0, state, q), c_n : (x_i, succ, x_j), c_n : (x_0, \\
 & conSucc, x_1) \rightarrow c_n : (x_1, head, x_j), c_n : (x_1, \sigma', x_i), c_n : (x_1, state, q_j)
 \end{aligned}$$

*Inertia* If in any configuration the head is at cell  $i$  of the tape, then in every successor configuration, elements in preceding and following cells of  $i$  in the tape are retained. The following two BRs ensures this:

$$\begin{aligned} & c_n : (x_0, head, x_i), c_n : (x_0, conSucc, x_1), c_n : (x_j, succt, x_i), c_n : (x_0, \sigma, x_j) \\ & \rightarrow c_n : (x_1, \sigma, x_j) \\ & c_n : (x_0, head, x_i), c_n : (x_0, conSucc, x_1), c_n : (x_i, succt, x_j), c_n : (x_0, \sigma, x_j) \\ & \rightarrow c_n : (x_1, \sigma, x_j) \end{aligned}$$

The rules above are instantiated for every  $\sigma \in \Sigma$ .

*Acceptance* A configuration whose state is  $q_A$  is accepting:

$$c_n : (x_0, state, q_A) \rightarrow c_n : (x_0, rdf : type, Accept)$$

If a configuration of accepting type is reached, then it can be back propagated to the initial configuration, using the following BR:

$$c_n : (x_0, conSucc, x_1), c_n : (x_1, rdf : type, Accept) \rightarrow c_n : (x_0, rdf : type, Accept)$$

Finally  $M$  accepts  $w$  iff the initial configuration is an accepting configuration. Let  $CQ^M$  be CCQ:  $\exists y \ c_n : (y, rdf : type, conInit), c_n : (y, rdf : type, Accept)$ . It can easily be verified that  $QS_C^M \models CQ^M$  iff the initial configuration is an accepting configuration. In order to prove the soundness and completeness of our simulation, we prove the following claims:

*Claim.* (1) The quad-system  $QS_C^M$  in the aforementioned simulation is a csafe quad-system

It can be noted that the only BRs in which existentials are present are the BRs used to generate the double exponential chain of tape cells and configurations, and are of the form (eBr). Note that in each of application of such a BR, a blank-node  $\_ : b$  generated in a context  $c_i$ , for any  $i = 1, \dots, n$ , is such that  $originContexts(\_ : b) = \{c_i\}$  and has exactly two child blank-nodes, each of whose origin contexts is  $\{c_{i-1}\}$ . Hence, any Skolem blank-node generated in any  $c_i$ , for  $i = 1 \dots n$  is such that its child blank-nodes has origin contexts  $c_{i-1}$ . Thanks to the above property, it turns out that there exists no two blank-nodes  $\_ : b, \_ : b'$  in the dChase of  $QS_C^M$  such that  $\_ : b$  is a descendant of  $\_ : b'$  and  $originContexts(\_ : b) = originContexts(\_ : b')$ . Therefore  $QS_C^M$  is csafe.

*Claim.* (2)  $QS_C^M \models CQ^M$  iff  $M$  accepts  $w$ .

Suppose that  $QS_C^M \models CQ^M$ , then by Theorem 1, there exists an assignment  $\mu : V(CQ^M) \rightarrow C$ , with  $CQ^M[\mu] \subseteq dChase(QS_C)$ . This implies that there exists a constant  $o$  in  $C(dChase(QS_C))$ , with  $\{c_n : (o, rdf : type, Accept), c_n : (o, rdf : type, conInit)\} \subseteq dChase(QS_C)$ . But thanks to the acceptance axioms it follows that there exists an constant  $o'$  such that  $\{c_n : (o, conSucc, o_1), c_n : (o_1, conSucc, o_2), \dots, c_n : (o_n, conSucc, o')\} \subseteq dChase(QS_C)$ , and  $c_n : (o', rdf : type, Accept) \in dChase(QS_C)$ .



Also thanks to the initialization axioms, it can be seen that  $o$  represents the initial configuration of  $M$  i.e. it represents the configuration in which the initial state is  $q_0$ , and the left end of the read-write tape contains  $w$  followed by trailing  $\square$ s, with the read-write head positioned at the first cell of the tape. Also the transition axioms makes sure that if  $c_n: (o, \text{conSucc}, o'') \in dChase(QS_C)$ , then  $o''$  represents a successor configuration of  $o$ . That is, if  $o$  represents the configuration in which  $M$  is at state  $q$  with read-write head at position  $pos$  of the tape that contains a letter  $\sigma \in \Sigma$ , and if  $\Delta(q, \sigma) = (q', \sigma', D)$ , then  $o''$  represents the configuration in which  $M$  is at state  $q'$ , in which read-write head is at the position  $pos - 1 / pos + 1$  depending on whether  $D = -1 / +1$ , and  $\sigma'$  is at the position  $pos$  of the tape. As a consequence of the above arguments, it follows that  $o'$  represents an accepting configuration of  $M$ , i.e. a configuration in which the state is  $q_A$ , the lone accepting, halting state. This means that  $M$  accepts the string  $w$ .

For the converse, we briefly show that if  $QS_C^M \not\models CQ^M$  then  $M$  does not accept  $w$ . Suppose that  $QS_C^M \not\models CQ^M$ , then by Theorem 1, for every assignment  $\mu: \mathbf{V}(CQ^M) \rightarrow \mathbf{C}$ , it should be the case that  $CQ^M[\mu] \not\subseteq dChase(QS_C)$ . By the initialization axioms, we know that there exists a constant  $o \in \mathbf{C}(dChase(QS_C))$  with  $c_n: (o, \text{rdf:type}, \text{conInit}) \in dChase(QS_C)$ . We know that  $o$  represents the initial configuration of  $M$ . Also by the initial construction axioms of  $QS_C^M$ , we know that  $o$  is the initial element of a double exponential chain of objects that are linearly ordered by property symbol  $\text{conSucc}$ . From transition axioms we know that, if, for any  $o''$ ,  $c_n: (o, \text{conSucc}, o'') \in dChase(QS_C)$ , then  $o''$  represents a valid successor configuration of  $o$ , which itself holds for  $o''$ , and so on. This means that for none of the succeeding double exponential configurations of  $M$ , the accepting state  $q_A$  holds. This means that  $M$  does not reach an accepting configuration with string  $w$ , and hence rejects it.

Since the construction above shows the existence of a polynomial time reduction of the word problem of a 2EXPTIME DTM, which is a 2EXPTIME-hard problem, to the CCQ entailment problem over csafe quad-systems, it immediately follows that CCQ entailment over csafe/msafe/safe quad-systems is 2EXPTIME-hard.

## 4.2 Procedure for detecting safe/msafe/safe quad-systems

In this subsection, we present a procedure for deciding whether a given quad-system is safe (resp. msafe, resp. csafe) or not. If the quad-system is safe (resp. msafe, resp. csafe), the result of the procedure is a *safe dChase* (resp. *msafe dChase*, *csafe dChase*) that contains the standard dChase, and can be used for query answering. Since the safety (resp. msafety, resp. csafety) property of a quad-system is attributed to the dChase of the quad-system, the procedure nevertheless performs the standard operations for computing the dChase, but also generate quads that indicate origin ruleIds and origin vectors (resp. origin ruleIds, resp. origin-contexts) of each Skolem blank node generated. In each iteration, a test for safety is performed, by checking the presence of Skolem blank-nodes that violate the safety (resp. msafety, resp. csafety) condition. In case a violation is detected, a distinguished quad is generated and the safe (resp. msafe, resp. csafe) dChase construction is aborted, prematurely. On the contrary, if there exists an iteration in which no new quad is generated, the safe (resp. msafe, resp. csafe) dChase computation stops with a completed safe (resp. msafe, resp. csafe) dChase that contains

the standard dChase. Since all the additional quads produced for accounting information use a distinguished context identifier  $c_c \notin \mathcal{C}$ , the computed safe (resp. msafe, resp. csafe) dChase itself can be used for standard query answering. Before getting to the details of the procedure, we give a few necessary definitions.

**Definition 7 (Context Scope).** *The context scope of a term  $t$  in a set of quad-patterns  $Q$ , denoted by  $cScope(t, Q)$  is given as:  $cScope(t, Q) = \{c \mid c: (s, p, o) \in Q, s = t \vee p = t \vee o = t\}$ .*

For any quad-system  $QS_C = \langle Q_C, R \rangle$ , let  $c_c$  be an ad hoc context identifier such that  $c_c \notin \mathcal{C}$ , then for  $r_i = body(r_i)(\mathbf{x}, \mathbf{z}) \rightarrow head(r_i)(\mathbf{x}, \mathbf{y}) \in R$ , we define transformations  $augS(r_i)$ ,  $augM(r_i)$ ,  $augC(r_i)$  as follows:

$$augS(r_i) = body(r_i)(\mathbf{x}, \mathbf{z}) \rightarrow head(r_i)(\mathbf{x}, \mathbf{y}) \wedge \forall y_j \in \{\mathbf{y}\} [ \bigwedge_{x_k \in \{\mathbf{x}\}} c_c: (x_k, descendantOf, y_j) \wedge c_c: (y_j, descendantOf, y_j) \wedge c_c: (y_j, originRuleId, i) \wedge c_c: (y_j, originVector, \mathbf{x}) ]$$

It should be noted that  $c_c: (y_j, originVector, \mathbf{x})$  is not a valid quad pattern, and is only used for notation brevity. In the actual implementation, vectors can be stored using an rdf container data structure such as `rdf:List`, `rdf:Seq` or by typecasting it as a string.

$$augM(r_i) = body(r_i)(\mathbf{x}, \mathbf{z}) \rightarrow head(r_i)(\mathbf{x}, \mathbf{y}) \wedge \forall y_j \in \{\mathbf{y}\} [ \bigwedge_{x_k \in \{\mathbf{x}\}} c_c: (x_k, descendantOf, y_j) \wedge c_c: (y_j, descendantOf, y_j) \wedge c_c: (y_j, originRuleId, i) ]$$

$$augC(r_i) = body(r_i)(\mathbf{x}, \mathbf{z}) \rightarrow head(r_i)(\mathbf{x}, \mathbf{y}) \wedge \forall y_j \in \{\mathbf{y}\} [ \bigwedge_{x_k \in \{\mathbf{x}\}} c_c: (x_k, descendantOf, y_j) \wedge c_c: (y_j, descendantOf, y_j) \wedge \bigwedge_{c \in cScope(y_j, head(r_i))} c_c: (y_j, originContext, c) ]$$

Intuitively, the transformation  $augS/augM/augC$  on a BR  $r_i$ , augments the head part of  $r_i$  with additional types of quad patterns, which are the following:

1.  $c_c: (x_k, descendantOf, y_j)$ , for every existentially quantified variable  $y_j$  in  $\mathbf{y}$  and universally quantified variable  $x_k \in \{\mathbf{x}\}$ . This is done because, during dChase computation any application of an assignment  $\mu$  to  $r_i$  such that  $\mathbf{x}[\mu] = \mathbf{a}$ , resulting in the generation of a Skolem blank node  $\_ : b = \mu^{ext(\mathbf{y})}(y_j)$ , any  $a_i \in \{\mathbf{a}\}$  is a descendant of  $\_ : b$ . Hence, due to these additional quad-patterns, quads of the form  $c_c: (a_i, descendantOf, \_ : b)$  are also produced, and in this way, keeps track of the descendants of any Skolem blank node produced.
2.  $c_c: (y_j, descendantOf, y_j)$ , in order to maintain also the reflexivity of ‘descendantOf’ relation.

3.  $c_c: (y_j, \text{originContext}, c)$ , for every existentially quantified variable  $y_j$  in  $\{\mathbf{y}\}$ , every  $c \in cScope(y_j, \text{head}(r_i))$ . This is done because during dChase computation, any application of an assignment  $\mu$  on  $r_i$ , such that  $\mathbf{x}[\mu] = \mathbf{a}$ , resulting in the generation of a Skolem blank node  $\_ : b = \mu^{ext(\mathbf{y})}(y_j)$ ,  $c$  is an origin context of  $\_ : b$ . Hence due to these additional quad-patterns, quads of the form  $c_c: (\_ : b, \text{originContext}, c)$  is also produced. In this way, we keep track of the origin-contexts of any Skolem blank node produced.
4.  $c_c: (y_j, \text{originVector}, \mathbf{a})$ , This is done because during the dChase computation, for any application of an assignment  $\mu$  on  $r_i$ , such that  $\mathbf{x}[\mu] = \mathbf{a}$ , resulting in the generation of a Skolem blank node  $\_ : b = \mu^{ext(\mathbf{y})}(y_j)$ ,  $\mathbf{a}$  is the origin vector of  $\_ : b$ . Hence, due to these additional quad-patterns, quads of the form  $c_c: (\_ : b, \text{originVector}, \mathbf{a})$  is also produced. In this way, we keep track of the origin vector of any Skolem blank node produced.
5.  $c_c: (y_j, \text{originRuleId}, i)$ , for every existentially quantified variable  $y_j$  in  $\{\mathbf{y}\}$ , in-order to keep track of the ruleId of the BR used to create any Skolem blank node.

It can be noticed that for any BR  $r_i$  without existentially quantified variables, the transformations  $augS/augM/augC$  leaves  $r_i$  unchanged. For any set of BRs  $R$ , let

$$\begin{aligned}
 augS(R) \text{ (resp. } augM(R), \text{ resp. } augC(R)) &= \bigcup_{r_i \in R} augS(r_i) \text{ (resp. } augM(r_i), \\
 \text{resp. } augC(r_i)) &\cup \{c_c: (x_1, \text{descendantOf}, z_1) \wedge c_c: (z_1, \text{descendantOf}, x_2) \\
 &\rightarrow c_c: (x_1, \text{descendantOf}, x_2)\}
 \end{aligned}$$

The function **unsafeTest** (resp. **unMSafeTest**, resp. **unCSafeTest**) defined below, given a BR  $r_i = \text{body}(r_i)(\mathbf{x}, \mathbf{z}) \rightarrow \text{head}(r_i)(\mathbf{x}, \mathbf{y})$ , an assignment  $\mu$ , and a quad-graph  $Q$  checks, if application of  $\mu$  on  $r_i$  violates the safety (resp. msafety, resp. csafety) condition on  $Q$ .

**unsafeTest**( $r_i, \mu, Q$ )=True iff  $\exists \_ : b, \_ : b' \in \mathbf{B}$ , with all the following conditions being satisfied:

- $\_ : b \in \{\mathbf{x}[\mu]\}$ , and
- $c_c: (\_ : b', \text{descendantOf}, \_ : b) \in Q$ , and
- $c_c: (\_ : b', \text{originRuleId}, i) \in Q$ , and
- $c_c: (\_ : b', \text{originVector}, \mathbf{a}) \in Q$ , and  $\mathbf{a} \cong \mathbf{x}[\mu]$ .

Intuitively, **unsafeTest** returns True, if  $\mu$  applied to  $r_i$  will produce a fresh Skolem blank node  $\_ : b''$ , whose child  $\_ : b \in \{\mathbf{x}[\mu]\}$ , and according to knowledge in  $Q$ ,  $\_ : b'$  is a descendant of  $\_ : b$  such that the origin ruleId of  $\_ : b'$  is  $i$  (which is also the origin ruleId of  $\_ : b''$ ) and the origin vector of  $\_ : b'$  is isomorphic to the origin vector of  $\mathbf{x}[\mu]$  (which is also the origin vector of  $\_ : b''$ ). The functions **unMSafeTest** and **unCSafeTest** are similarly defined as follows:

**unMSafeTest**( $r_i, \mu, Q$ )=True iff  $\exists \_ : b, \_ : b' \in \mathbf{B}$ , with all the following conditions being satisfied:

- $\_ : b \in \{\mathbf{x}[\mu]\}$ , and
- $c_c: (\_ : b', \text{descendantOf}, \_ : b) \in Q$ , and

- $c_c: (\_ : b', \text{originRuleId}, i) \in Q$ .

**unCSafeTest**( $r_i, \mu, Q$ )=True iff  $\exists \_ : b, \_ : b' \in \mathbf{B}, \exists y_j \in \{\mathbf{y}\}$ , with all the following being satisfied:

- $\_ : b \in \{\mathbf{x}[\mu]\}$ , and
- $c_c: (\_ : b', \text{descendantOf}, \_ : b) \in Q$ , and
- $\{c \mid c_c: (\_ : b', \text{originContext}, c) \in Q\} = cScope(y_j, \text{head}(r_i)(\mathbf{x}, \mathbf{y})) \setminus \{c_c\}$ .

For any BR  $r_i$  and an assignment  $\mu$ , the *safe/msafe/csaf*e application of  $\mu$  on  $r_i$  w.r.t. a quad-graph  $Q_C$  is defined as follows:

$$\begin{aligned} \text{apply}^{\text{safe}}(r_i, \mu, Q_C) &= \begin{cases} \mathbf{unSafe}, & \text{If } \text{unSafeTest}(r_i, \mu, Q_C) = \text{True}; \\ \text{apply}(r_i, \mu), & \text{Otherwise}; \end{cases} \\ \text{apply}^{\text{msafe}}(r_i, \mu, Q_C) &= \begin{cases} \mathbf{unMSafe}, & \text{If } \text{unMSafeTest}(r_i, \mu, Q_C) = \text{True}; \\ \text{apply}(r_i, \mu), & \text{Otherwise}; \end{cases} \\ \text{apply}^{\text{csafe}}(r_i, \mu, Q_C) &= \begin{cases} \mathbf{unCSafe}, & \text{If } \text{unCSafeTest}(r_i, \mu, Q_C) = \text{True}; \\ \text{apply}(r_i, \mu), & \text{Otherwise}; \end{cases} \end{aligned}$$

where **unSafe** =  $c_c: (\text{unsafe}, \text{unsafe}, \text{unsafe})$  (resp. **unMSafe** =  $c_c: (\text{unmsafe}, \text{unmsafe}, \text{unmsafe})$ , resp. **unCSafe** =  $c_c: (\text{uncsafe}, \text{uncsafe}, \text{uncsafe})$ ) is a distinguished quad that is generated, if the prerequisites of safety (resp. msafety, resp. csafety) is violated. For any quad-system  $Q_{SC} = \langle Q_C, R \rangle$ , we define its *safe dChase*  $dChase^{\text{safe}}(Q_{SC})$  as follows:

$$\begin{aligned} dChase_0^{\text{safe}}(Q_{SC}) &= Q_C; dChase_{m+1}^{\text{safe}}(Q_{SC}) = dChase_m^{\text{safe}}(Q_{SC}) \cup \text{apply}^{\text{safe}}(r_i, \mu, dChase_m^{\text{safe}}(Q_{SC})), \text{ if } \exists r_i \in \text{augS}(R), \text{ assignment } \mu \text{ such that } \text{applicable}_{\text{augS}(R)}(r_i, \mu, dChase_m^{\text{safe}}(Q_{SC})); \\ dChase_{m+1}^{\text{safe}}(Q_{SC}) &= dChase_m^{\text{safe}}(Q_{SC}), \text{ otherwise; for any } m \in \mathbb{N}. \\ dChase^{\text{safe}}(Q_{SC}) &= \bigcup_{m \in \mathbb{N}} dChase_m^{\text{safe}}(Q_{SC}) \end{aligned}$$

The termination condition for safe dChase computation can be implemented using the following conditional: If there exists  $m$  such that

$$\begin{aligned} dChase_m^{\text{safe}}(Q_{SC}) &= dChase_{m+1}^{\text{safe}}(Q_{SC}); \text{ then} \\ dChase^{\text{safe}}(Q_{SC}) &= dChase_m^{\text{safe}}(Q_{SC}). \end{aligned}$$

The dChases  $dChase^{\text{msafe}}(Q_{SC})$  and  $dChase^{\text{csafe}}(Q_{SC})$  are defined, similarly, for msafe and csafe quad-systems, respectively. We bring to the notice of the reader that although application of any  $\text{augS}(r)$  (resp.  $\text{augM}(r)$ , resp.  $\text{augC}(r)$ ) produces quad-patterns of the form  $c_c: (\_ : b, \text{descendantOf}, \_ : b)$ , for any Skolem blank node  $\_ : b$  generated, there is no raise of a false alarm in the unSafeTest (resp. unMSafeTest, resp. unCSafeTest). This is because unSafeTest (resp. unMSafeTest, resp. unCSafeTest) on a bridge rule  $r = \text{body}(r)(\mathbf{x}, \mathbf{z}) \rightarrow \text{head}(r)(\mathbf{x}, \mathbf{y})$  and assignment  $\mu$  checks if the application of  $\mu$  of  $r$  with the fresh  $\_ : b''$  assigned to a  $y_i \in \{\mathbf{y}\}$  by  $\mu^{\text{ext}(\mathbf{y})}$  would have a child  $\_ : b \neq b''$  assigned to some  $x_i \in \{\mathbf{x}\}$  by  $\mu$ , such that there exists a quad of the form  $c_c: (\_ : b', \text{descendantOf}, \_ : b)$  in the safe (resp. msafe, resp. csafe) dChase constructed so far, and  $\_ : b''$  and  $\_ : b'$  have the same origin ruleId and originVector (resp. originRuleId, resp. originContexts). Note that in the above  $\_ : b'$  should also be distinct from  $\_ : b''$ , and hence rules out the case in which unSafeTest (resp. unMSafeTest, resp. unCSafeTest) returns True because of the detection of a blank node as a self descendant of itself.

The following theorem shows that the procedure above described for detecting unsafe quad-systems is sound and complete:

**Theorem 5.** *For any quad-system  $QS_C = \langle Q_C, R \rangle$ , the quad **unSafe** (resp. **unMSafe**, resp. **unCSafe**)  $\in dChase^{safe}(QS_C)$  (resp.  $dChase^{msafe}(QS_C)$ , resp.  $dChase^{csafe}(QS_C)$ ), iff  $QS_C$  is unsafe (resp. unmsafe, resp. uncsafe).*

It should be noted that for any quad-system  $QS_C = \langle Q_C, R \rangle$ ,  $dChase^{safe}(QS_C)$  (resp.  $dChase^{msafe}(QS_C)$ , resp.  $dChase^{csafe}(QS_C)$ ) is a finite set and hence the iterative procedure which we described earlier terminates, regardless of whether  $QS_C$  is safe (resp. msafe, resp. csafe) or not. This is because if  $QS_C$  is safe (resp. msafe, resp. csafe), then, as we have seen before, there exists a double exponential bound on number of quads in its dChase. Hence, there is an iteration in which no new quad is generated, which leads to stopping of computation. Otherwise, if  $QS_C$  is unsafe (resp. msafe, resp. csafe), then from Theorem 5, we know that the quad **unSafe** (resp. **unMSafe**, resp. **unCSafe**) gets generated in  $dChase^{safe}(QS_C)$  (resp.  $dChase^{msafe}(QS_C)$ , resp.  $dChase^{csafe}(QS_C)$ ) in not more than  $\mathcal{O}(2^{2^{\|QS_C\|}})$  iterations. This implies that there exists an iteration  $m$  such that the quad **unSafe** (resp. **unMSafe**, resp. **unCSafe**) is in  $dChase_m^{safe}(QS_C)$  (resp.  $dChase_m^{msafe}(QS_C)$ , resp.  $dChase_m^{csafe}(QS_C)$ ). W.l.o.g, let  $m$  be the first such iteration. This means that there exists a BR  $r_i \in R$  with head  $head(r_i)(x, y)$ , assignment  $\mu$  such that  $applicable_{augS(R)}(r_i, \mu, dChase_{m-1}^{safe}(QS_C))$  (resp.  $applicable_{augM(R)}(r_i, \mu, dChase_{m-1}^{msafe}(QS_C))$ , resp.  $applicable_{augC(R)}(r_i, \mu, dChase_{m-1}^{csafe}(QS_C))$ ) holds. By construction, since  $head(r_i)[\mu^{ext(y)}]$  is not generated, and instead the quad **unSafe** (resp. **unMSafe**, resp. **unCSafe**) is generated,  $applicable_{augS(R)}(r_i, \mu, dChase_m^{safe}(QS_C))$  (resp.  $applicable_{augM(R)}(r_i, \mu, dChase_m^{msafe}(QS_C))$ , resp.  $applicable_{augC(R)}(r_i, \mu, dChase_m^{csafe}(QS_C))$ ) holds yet again. This means that the termination condition is satisfied at iteration  $m + 1$ , and hence computation stops. Note that regardless of whether a given quad-system is safe (resp. msafe, resp. csafe) or not, the number of safe (resp. msafe, resp. csafe) dChase iterations is double exponentially bounded in the size of the quad-system. Consequently, we derive the following theorem.

**Theorem 6.** *Recognizing whether a quad-system is safe/msafe/csafe is in 2EXPTIME.*

Also notice that after running procedure described above, if the quad **unSafe** (resp. **unMSafe**, resp. **unCSafe**) is not generated, then its safe (resp. msafe, resp. csafe) dChase itself can be used for CCQ answering, as in such a case the standard dChase is contained in safe (resp. msafe, resp. csafe) dChase, and all the quads generated for accounting information have the context identifier  $c_c$ . Hence, for any safe (resp. msafe, resp. csafe) quad-system, for any boolean CCQ that does not contain quad patterns of the form  $c_c: (s, p, o)$ , the dChase entails CCQ iff the safe (resp. msafe, resp. csafe) dChase entails CCQ.

A set of BRs  $R$  is said to be *universally safe* (resp. *msafe*, resp. *csafe*) iff, for any quad-graph  $Q_C$ , the quad-system  $\langle Q_C, R \rangle$  is safe (resp. msafe, resp. csafe). For any set of BRs  $R$ , whose set of context identifiers is  $\mathcal{C}$ , also let  $U_R$  be the set of URIs that occur in the triple patterns of  $R$  plus an additional ad hoc blank node  $\_ : b^{crit}$ , the *critical quad-graph* of  $R$  is defined as the set  $\{c: (s, p, o) | c \in \mathcal{C}, \{s, p, o\} \subseteq U_R\}$ . The following property illustrates how the critical quad-graph of a set of BRs  $R$  can be used to determine, whether or not  $R$  is universally safe/msafe/csafe.

*Property 4.* A set of BRs  $R$  is universally safe (resp. msafe, resp. csafe) iff  $\langle Q_C^{crit}, R \rangle$  is safe (resp. msafe, resp. csafe), where  $Q_C^{crit}$  is the critical quad-graph of  $R$ .

## 5 Range Restricted Quad-Systems: Restricting to Range Restricted BRs

In this section, we investigate the complexity of CCQ entailment over quad-systems, whose BRs do not have existentially quantified variables. Such BRs are of the form:

$$c_1 : t_1(\mathbf{x}, \mathbf{z}) \wedge \dots \wedge c_n : t_n(\mathbf{x}, \mathbf{z}) \rightarrow c'_1 : t'_1(\mathbf{x}) \wedge \dots \wedge c'_m : t'_m(\mathbf{x})$$

Note that any set of BRs  $R$  of the form above can be replaced by semantically equivalent set  $R'$ , such that each  $r \in R'$  is the form:

$$c_1 : t_1(\mathbf{x}, \mathbf{z}), \dots, c_n : t_n(\mathbf{x}, \mathbf{z}) \rightarrow c'_1 : t'_1(\mathbf{x}) \quad (5)$$

Also  $\|R'\|$  is at most quadratic in  $\|R\|$ , and hence, w.l.o.g, we assume that each  $r \in R$  is of the form (5). Borrowing the parlance from the  $\forall\exists$  rules setting, where rules whose variables in the head part are contained in the variables in the body part are called range restricted rules [15], we call such BRs *range restricted* (RR) BRs. We call a quad-system whose BRs are all of RR-type, a *RR quad-system*. Since there exists no existentially quantified variable in the BRs of a RR quad-system, no Skolem blank node is produced during dChase computation. Hence, there can be no violation of the safety/msafety/csafety condition in section 4, and hence, the class of RR quad-systems are contained in the class of safe/msafe/csafes quad-systems, and is also a FEC. Of course, this containment is strict as any quad-system that contains a BR with an existential variable is not RR. Since one can determine whether or not a given quad-system is RR or not by simply iterating through set of BRs and checking their syntax, the following holds:

**Theorem 7.** *Recognizing whether a quad-system is RR can be done in linear time.*

In the following, we see that restricting to RR BRs, size of the dChase becomes polynomial w.r.t. size of the input quad-system, and the complexity of CCQ entailment further reduces compared to safe/msafe/csafes quad-systems.

**Lemma 5.** *For any RR quad-system  $QS_C = \langle Q_C, R \rangle$ , the following holds: (i)  $\|dChase(QS_C)\| = \mathcal{O}(\|QS_C\|^4)$  (ii)  $dChase(QS_C)$  can be computed in EXPTIME (iii) If  $\|R\|$  is fixed to be a constant,  $dChase(QS_C)$  can be computed in PTIME.*

*Proof.* (i) Note that the number of constants in  $QS_C$  is roughly equal to  $\|QS_C\|$ . As no existential variable occurs in any BR in a RR quad-system  $QS_C$ , the set of constants  $\mathbf{C}(dChase(QS_C))$  is contained in  $\mathbf{C}(QS_C)$ . Since each  $c: (s, p, o) \in dChase(QS_C)$  is such that  $c, s, p, o \in \mathbf{C}(QS_C)$ ,  $|dChase(QS_C)| = \mathcal{O}(|\mathbf{C}(QS_C)|^4)$ . Hence  $\|dChase(QS_C)\| = \mathcal{O}(|\mathbf{C}(QS_C)|^4) = \mathcal{O}(\|QS_C\|^4)$ .

(ii) Since from (i)  $|dChase(QS_C)| = \mathcal{O}(\|QS_C\|^4)$ , and in each iteration of the dChase at least one new quad is added, the number of iterations cannot exceed  $\mathcal{O}(\|QS_C\|^4)$ .

Since by Lemma 1, computation of each iteration  $i$  of the dChase requires  $\mathcal{O}(|R| * \|dChase_{i-1}(QS_C)\|^{rs})$  time, where  $rs = \max_{r \in R} \|r\|$ , and  $rs \leq \|QS_C\|$ , time required for each iteration is of the order  $\mathcal{O}(2^{\|QS_C\|})$  time. Although the number of iterations is a polynomial, each iteration requires an exponential amount of time w.r.t  $\|QS_C\|$ . Hence time complexity of dChase computation is in EXPTIME.

(iii) As we know that the time taken for application of a BR  $R$  is  $\mathcal{O}(\|dChase_{i-1}(QS_C)\|^{\|R\|})$ . Since  $\|R\|$  is fixed to a constant, application of  $R$  can be done in PTIME. Hence, each dChase iteration can be computed in PTIME. Also since the number of iterations is a polynomial in  $\|QS_C\|$ , computing dChase is in PTIME.

**Theorem 8.** *Data complexity of CCQ entailment over RR quad-systems is PTIME-complete.*

*Proof.* (Membership) Follows from the membership in P of data complexity of CCQ entailment for safe quad-systems, whose expressivity subsumes the expressivity of RR quad-systems (Theorem 4).

(Hardness) In order to prove P-hardness, we reduce a well known P-complete problem, 3HornSat, i.e. the satisfiability of propositional Horn formulas with at most 3 literals. Note that a (propositional) Horn formula is a propositional formula of the form:

$$P_1 \wedge \dots \wedge P_n \rightarrow P_{n+1} \quad (6)$$

where  $P_i$ , for  $1 \leq i \leq n+1$ , are either propositional variables or constants  $t, f$ , that represents true and false, respectively. Note that for any propositional variable  $P$ , the fact that “ $P$  holds” is represented by the formula  $t \rightarrow P$ , and “ $P$  does not hold” is represented by the formula  $P \rightarrow f$ . A 3Horn formula is a formula of the form (6), where  $1 \leq n \leq 2$ . Note that any (set of) Horn formula(s)  $\Phi$  can be transformed in polynomial time to a polynomially sized set  $\Phi'$  of 3Horn formulas, by introducing auxiliary propositional variables such that  $\Phi$  is satisfiable iff  $\Phi'$  is satisfiable. A pure 3Horn formula is a 3Horn formula of the form (6), where  $n = 2$ . Any 3Horn formula  $\phi$  that is not pure can be trivially converted to equivalent pure form by appending a  $\wedge t$  on the body part of  $\phi$ . For instance,  $P \rightarrow Q$ , can be converted to  $P \wedge t \rightarrow Q$ . Hence, w.l.o.g. we assume that any set of 3Horn formulas is pure, and is of the form:

$$P_1 \wedge P_2 \rightarrow P_3 \quad (7)$$

In the following, we reduce the satisfiability problem of pure 3Horn formulas to CCQ entailment problem over a quad-system whose set of schema triples, the set of BRs, and the CCQ  $CQ$  are all fixed.

For any set of pure Horn formulas  $\Phi$ , we construct the quad-system  $QS_C = \langle Q_C, R \rangle$ , where  $C = \{c_t, c_f\}$ . For any formula  $\phi \in \Phi$  of the form (7),  $Q_C$  contains a quad  $c_f: (P_1, P_2, P_3)$ . In addition  $Q_C$  contains a quad  $c_t: (t, \text{rdf} : \text{type}, T)$ .  $R$  is the singleton that contains only the following fixed BR:

$$c_t: (x_1, \text{rdf} : \text{type}, T), c_t: (x_2, \text{rdf} : \text{type}, T), c_f: (x_1, x_2, x_3) \rightarrow c_t: (x_3, \text{rdf} : \text{type}, T)$$

Let the  $CQ$  be the fixed query  $c_t: (f, \text{rdf} : \text{type}, T)$ .

Now, it is easy to see that  $QS_C \models CQ$ , iff  $\Phi$  is not satisfiable.

**Theorem 9.** *Combined complexity of CCQ entailment over RR quad-systems is in EXPTIME.*

*Proof.* (Membership) By Lemma 5, for any RR quad-system  $QS_C$ , its dChase  $dChase(QS_C)$  can be computed in EXPTIME. Also by Lemma 5, its dChase size  $\|dChase(QS_C)\|$  is a polynomial w.r.t to  $\|QS_C\|$ . A boolean CCQ  $CQ()$  can naively be evaluated by grounding the set of constants in the dChase to the variables in the  $CQ()$ , and then checking if any of these groundings are contained in  $dChase(QS_C)$ . The number of such groundings can at most be  $\|dChase(QS_C)\|^{\|CQ()\|}$  ( $\dagger$ ). Since  $\|dChase(QS_C)\|$  is a polynomial in  $\|QS_C\|$ , there are an exponential number of groundings w.r.t  $\|CQ()\|$ . Since containment of each of these groundings can be checked in time polynomial w.r.t. the size of  $dChase(QS_C)$ , and since  $\|dChase(QS_C)\|$  is a polynomial w.r.t.  $\|QS_C\|$ , the time complexity of CCQ entailment is in EXPTIME.

Concerning the combined complexity of CCQ entailment of RR quad-systems, we leave the lower bounds open.

### 5.1 Restricted RR Quad-Systems

We call those quad-systems with BRs of form (5) with a fixed bound on  $n$  as *restricted RR quad-systems*. They can be further classified as linear, quadratic, cubic,..., quad-systems, when  $n = 1, 2, 3, \dots$ , respectively.

**Theorem 10.** *Data complexity of CCQ entailment over restricted RR quad-systems is P-complete.*

*Proof.* The proof is same as in Theorem 8, since the size of BRs are fixed to constant.

**Theorem 11.** *Combined complexity of CCQ entailment over restricted RR quad-systems is NP-complete.*

*Proof.* Let the problem of deciding if  $QS_C \models CQ()$  be called DP'.

(Membership) for any  $QS_C$  whose rules are of restricted RR-type, the size of any  $r \in R$  is a constant. Hence, by Lemma 1, any dChase iteration can be computed in PTIME. Since the number of iterations is also polynomial in  $\|QS_C\|$ ,  $dChase(QS_C)$  can be computed in PTIME in the size of  $QS_C$  and  $dChase(QS_C)$  has a polynomial number of constants. Hence, we can guess an assignment  $\mu$  for all the existential variables in CCQ  $CQ()$ , to the set of constants in  $dChase(QS_C)$ . Then, one can evaluate the CCQ, by checking if  $c: (s, p, o) \in dChase(QS_C)$ , for each  $c: (s, p, o) \in CQ()[\mu]$ , which can be done in time  $\mathcal{O}(\|CQ\| * \|dChase(QS_C)\|)$ , and is hence in non-deterministic PTIME, which implies that DP' is in NP.

(Hardness) We show that DP' is NP-hard, by reducing the well known NP-hard problem of 3-colorability to DP'. Given a graph  $G = \langle V, E \rangle$ , where  $V = \{v_1, \dots, v_n\}$  is the set of nodes,  $E \subseteq V \times V$  is the set of edges, the 3-colorability problem is to decide if there exists a labeling function  $l: V \rightarrow \{r, b, g\}$  that assigns each  $v \in V$  to an element in  $\{r, b, g\}$  such that the condition:  $(v, v') \in E \rightarrow l(v) \neq l(v')$ , for each  $(v, v') \in E$ , is satisfied.



One can construct a quad-system  $QS_c = \langle Q_c, \emptyset \rangle$ , where  $graph_{Q_c}(c)$  has the following triples:

$$\{(r, edge, b), (r, edge, g), (b, edge, g), (b, edge, r), (g, edge, r), (g, edge, b)\}$$

Let  $CQ$  be the boolean CCQ:  $\exists v_1, \dots, v_n \bigwedge_{(v, v') \in E} [c: (v, edge, v') \wedge c: (v', edge, v)]$ .

Then, it can be seen that  $G$  is 3-colorable, iff  $QS_c \models CQ$ .

## 6 Quad-Systems and Forall-Existential rules: A formal comparison

In this section, we formally compare the formalism of quad-systems with forall-existential ( $\forall\exists$ ) rules, which are also called Tuple generating dependencies (Tgds)/Datalog+- rules.  $\forall\exists$  rules is a fragment of first order logic in which every formula is restricted to a certain syntactic form. A  $\forall\exists$  rule is a first order formula of the form:

$$\forall \mathbf{x} \forall \mathbf{z} [p_1(\mathbf{x}, \mathbf{z}) \wedge \dots \wedge p_n(\mathbf{x}, \mathbf{z}) \rightarrow \exists \mathbf{y} p'_1(\mathbf{x}, \mathbf{y}) \wedge \dots \wedge p'_m(\mathbf{x}, \mathbf{y})] \quad (8)$$

where  $\mathbf{x}, \mathbf{y}, \mathbf{z}$  are vectors of variables such that  $\{\mathbf{x}\}, \{\mathbf{y}\}$  and  $\{\mathbf{z}\}$  are pairwise disjoint,  $p_i(\mathbf{x}, \mathbf{z})$ , for  $1 \leq i \leq n$  are predicate atoms whose variables are from  $\mathbf{x}$  or  $\mathbf{z}$ ,  $p'_i(\mathbf{x}, \mathbf{y})$ , for  $1 \leq i \leq m$  are predicate atoms whose variables are from  $\mathbf{x}$  or  $\mathbf{y}$ . We, for short, occasionally note a  $\forall\exists$  rule of the form (8) as  $\phi(\mathbf{x}, \mathbf{z}) \rightarrow \psi(\mathbf{x}, \mathbf{y})$ , where  $\phi(\mathbf{x}, \mathbf{z}) = \{p_1(\mathbf{x}, \mathbf{z}), \dots, p_n(\mathbf{x}, \mathbf{z})\}$ ,  $\psi(\mathbf{x}, \mathbf{y}) = \{p'_1(\mathbf{x}, \mathbf{y}), \dots, p'_m(\mathbf{x}, \mathbf{y})\}$ . A set of  $\forall\exists$  rules is called a  $\forall\exists$  rule set. In the realm of  $\forall\exists$  rule sets, a conjunctive query (CQ) is an expression of the form:

$$\exists \mathbf{y} p_1(\mathbf{x}, \mathbf{y}) \wedge \dots \wedge p_r(\mathbf{x}, \mathbf{y}) \quad (9)$$

where  $p_i(\mathbf{x}, \mathbf{y})$ , for  $1 \leq i \leq r$  are predicate atoms over vectors  $\mathbf{x}$  or  $\mathbf{y}$ . A boolean CQ is defined as usual. The decision problem of whether, for a  $\forall\exists$  rule set  $\mathbb{P}$  and a CQ  $Q$ , if  $\mathbb{P} \models_{\text{fol}} Q$  is called the  $CQ\ EP$ , where  $\models_{\text{fol}}$  is the standard first order logic entailment relation.

For any quad-graph  $Q_C = \{c_1: (s_1, p_1, o_1), \dots, c_n: (s_r, p_r, o_r)\}$ , let  $r_{Q_C}$  be the BR

$$\rightarrow \exists y_{b_1}, \dots, y_{b_q} c_1: (s_1, p_1, o_1)[\mu_{\mathbf{B}}] \wedge \dots \wedge c_r: (s_r, p_r, o_r)[\mu_{\mathbf{B}}],$$

where  $\{-: b_1, \dots, -: b_q\}$  is the set of blank nodes in  $Q_C$ , and  $\mu_{\mathbf{B}}$  is the substitution function  $\{-: b_i \rightarrow y_{b_i}\}_{i=1, \dots, q}$  that assigns each blank-node to a fresh existentially quantified variable. It can be noted that the quad-systems  $\langle Q_C, R \rangle$  and  $\langle \emptyset, R \cup \{r_{Q_C}\} \rangle$  are semantically equivalent. The following definition gives the translation functions that will be necessary to establish the relation between quad-systems and  $\forall\exists$  rule sets.

**Definition 8 (Translations  $\tau_q, \tau_r, \tau_{ccq}, \tau$ ).** *The translation function  $\tau_q$  from the set of quad patterns to the set of ternary atoms is defined as: for any quad-pattern  $c: (s, p, o)$ ,  $\tau_q(c: (s, p, o)) = c(s, p, o)$ .*

*The translation function  $\tau_{br}$  from the set of BRs to the set of  $\forall\exists$  rules is defined as: for any BR  $r$  of the form (2):*

$$\tau_{br}(r) = \forall \mathbf{x} \forall \mathbf{z} [\tau_q(c_1: t_1(\mathbf{x}, \mathbf{z})) \wedge \dots \wedge \tau_q(c_n: t_n(\mathbf{x}, \mathbf{z})) \rightarrow \exists \mathbf{y} \tau_q(c'_1: t'_1(\mathbf{x}, \mathbf{y})) \wedge \dots \wedge \tau_q(c'_m: t'_m(\mathbf{x}, \mathbf{y}))],$$

The translation function  $\tau$  from the set of quad-systems to forall-existential rule sets is defined as: for any quad-system  $QS_C = \langle Q_C, R \rangle$ ,  $\tau(QS_C) = \tau_{br}(R) \cup \{\tau_{br}(r_{Q_C})\}$ , where  $\tau_{br}(R) = \bigcup_{r \in R} \tau_{br}(r)$ .

The translation function  $\tau_{ccq}$  from the set of boolean CCQs to the set of boolean CQs is defined as: for any boolean CCQ  $CQ = \exists \mathbf{y} \ c_1 : t_1(\mathbf{a}, \mathbf{y}) \wedge \dots \wedge c_r : t_r(\mathbf{a}, \mathbf{y})$ ,  $\tau_{ccq}(CQ)$  is:

$$\exists \mathbf{y} \ \tau_q(c_1 : t_1(\mathbf{a}, \mathbf{y})) \wedge \dots \wedge \tau_q(c_r : t_r(\mathbf{a}, \mathbf{y})).$$

The following property gives the relation between CCQ entailment of unrestricted quad-systems and standard first order CQ entailment of  $\forall\exists$  rule sets.

*Property 5.* For any quad-system  $QS_C$ , CCQ  $CQ$ ,  $QS_C \models CQ$  iff  $\tau(QS_C) \models_{fol} \tau_{ccq}(CQ)$ .

*Proof.* Notice that every context  $c \in \mathcal{C}$  becomes a ternary predicate symbol in the resulting translation. Also,  $\tau(QS_C)$  is a  $\forall\exists$  rule set, and for any CCQ  $CQ$ ,  $\tau_{ccq}(CQ)$  is a CQ.

In order to construct the restricted chase for  $\tau(QS_C)$ , suppose that  $\prec_q$  is also extended to set of instances such that for any two quad-graphs  $Q_C, Q'_C$ ,  $Q_C \prec_q Q'_C$  iff  $\tau_q(Q_C) \prec_q \tau_q(Q'_C)$ . Suppose  $\prec$  is extended similarly to set of instances. Also assume that during the construction of standard chase  $chase(\tau(QS_C))$  of  $\tau(QS_C)$ , for any application of a  $\tau_{br}(r)$  with existential variables, with  $r \in R$ , suppose that the Skolem blank nodes generated in  $chase(\tau(QS_C))$  follow the same order as they are generated in  $dChase(QS_C)$ . Also let us extend the rule applicability function to the  $\forall\exists$  rules settings such that for any set of BRs  $R$ , for any  $r \in R$ , quad-graph  $Q'_C$ , assignment  $\mu$ ,  $applicable_R(r, \mu, Q'_C)$  iff  $applicable_{\tau_{br}(R)}(\tau_{br}(r), \mu, \tau_q(Q'_C))$ .

Now  $dChase_0(\langle \emptyset, R \cup \{r_{Q_C}\} \rangle) = \emptyset$ , and also  $chase_0(\tau(QS_C)) = \emptyset$ ,  $dChase_1(QS_C) = apply(r_{Q_C}, \mu_\emptyset)$ , where  $\mu_\emptyset$  is the empty function,  $chase_1(\tau(QS_C)) = apply(\tau_{br}(r_{Q_C}), \mu_\emptyset)$ , and so on. It is straightforward to see that for any  $m \in \mathbb{N}$ ,  $\tau_q(dChase_m(\langle \emptyset, R \cup \{r_{Q_C}\} \rangle)) = chase_m(\tau(QS_C))$ . As a consequence,  $\tau_q(dChase(QS_C)) = chase(\tau(QS_C))$ , and  $\{CQ\}[\sigma] \subseteq dChase(QS_C)$  iff  $\{\tau_{ccq}(CQ)\}[\sigma] \subseteq chase(\tau(QS_C))$ .

Hence, applying Theorem 1 and the analogous theorem for  $\forall\exists$  rulesets from Deutch et al. [30], it follows that for any quad-system  $QS_C = \langle Q_C, R \rangle$  and a boolean CCQ  $CQ$ ,  $QS_C \models CQ$  iff  $\tau(QS_C) \models_{fol} \tau_{ccq}(CQ)$ .

**Theorem 12.** *There exists a polynomial time translation function  $\tau$  (resp.  $\tau_{ccq}$ ) from the set of unrestricted quad-systems (resp. CCQs) to the set of  $\forall\exists$  rule sets (resp. CQs), such that for any unrestricted quad-system  $QS_C$  and a CCQ  $CQ$ ,  $QS_C \models CQ$  iff  $\tau(QS_C) \models_{fol} \tau_{ccq}(CQ)$ .*

*Proof.* It is easy to see that  $\tau_q$ ,  $\tau_{br}$ ,  $\tau$ , and  $\tau_{ccq}$  in Definition 8 can be implemented using simple syntax transformation, by iterating through the respective components of a quad-system/CCQ, and the time complexity of these functions are linear w.r.t their inputs.

Notice that for any CCQ  $CQ$  (resp. CQ  $Q$ ),  $\rightarrow CQ$  (resp.  $\rightarrow Q$ ) is a bridge (resp.  $\forall\exists$ ) rule, with an empty body. Also, since for any quad-graph  $Q_C$ , the translation function  $\tau_{br}$  defined above can directly be applied on  $r_{Q_C}$  to obtain a  $\forall\exists$  rule, the following theorem immediately follows:

**Theorem 13.** For quad-systems, the EPs: (i) quad EP, (ii) quad-graph EP, (iii) BR EP, (iv) BRs EP, (v) Quad-System EP, and (vi) CCQ EP are polynomially reducible to entailment of  $\forall\exists$  rule sets.

A  $\forall\exists$  rule set  $\mathbb{P}$  is said to be a *ternary*  $\forall\exists$  rule set, iff all the predicate symbols in the vocabulary of  $\mathbb{P}$  are of arity less than or equal to three.  $\mathbb{P}$  is a *purely ternary* rule set, iff all the predicate symbols in the vocabulary  $\mathbb{P}$  is of arity three. Similarly, a (purely) ternary CQ is defined. The following property gives the relation between the CQ entailment problem of  $\forall\exists$  rule sets and CCQ EP of unrestricted quad-systems.

**Theorem 14.** There exists a polynomial time translation function  $\nu$  (resp.  $\nu_{cq}$ ) from ternary  $\forall\exists$  rule sets (resp. ternary CQs) to unrestricted quad-systems (resp. CCQs) such that for any ternary  $\forall\exists$  rule set  $\mathbb{P}$  and a ternary CQ  $Q$ ,  $\mathbb{P} \models_{fol} CQ$  iff  $\langle \emptyset, \nu(\mathbb{P}) \rangle \models \nu_{cq}(Q)$ .

*Proof.* Note that the CQ EP of any ternary  $\forall\exists$  rule set  $\mathbb{P}$ , whose set of predicate symbols is  $P$ , and CQ  $Q$  over  $P$ , can polynomially reduced to the CQ EP of a purely ternary rule set  $\mathbb{P}'$  and purely ternary CQ  $Q'$ , by the following transformation function  $\chi$ . Let  $\square$  be an adhoc fresh URI;  $\chi$  is such that for any ternary atom  $c(s, p, o)$ ,  $\chi(c(s, p, o)) = c(s, p, o)$ . For any binary atom  $c(s, p)$ ,  $\chi(c(s, p)) = c(s, p, \square)$ , and for any unary atom  $c(s)$ ,  $\chi(c(s)) = c(s, \square, \square)$ . For any  $\forall\exists$  rule  $r$  of the form (8),

$$\begin{aligned} \chi(r) &= \forall \mathbf{x} \forall \mathbf{z} [\chi(p_1(\mathbf{x}, \mathbf{z})) \wedge \dots \wedge \chi(p_n(\mathbf{x}, \mathbf{z})) \\ &\rightarrow \exists \mathbf{y} \chi(p'_1(\mathbf{x}, \mathbf{y})) \wedge \dots \wedge \chi(p'_m(\mathbf{x}, \mathbf{y}))] \end{aligned}$$

And, for any  $\forall\exists$  rule set  $\mathbb{P}$ ,  $\chi(\mathbb{P}) = \bigcup_{r \in \mathbb{P}} \chi(r)$ . For any CQ  $Q$ ,  $\chi(Q)$  is similarly defined. Note that for any ternary  $\forall\exists$  rule set  $\mathbb{P}$ , ternary CQ  $Q$ ,  $\chi(\mathbb{P})$  (resp.  $\chi(Q)$ ) is purely ternary, and  $\mathbb{P} \models_{fol} Q$  iff  $\chi(\mathbb{P}) \models_{fol} \chi(Q)$ .

Also, it can straightforwardly seen that  $\tau_{br}^{-1}(\chi(\mathbb{P}))$  (resp.  $\tau_{ccq}^{-1}(\chi(Q))$ ) is a set of BRs (resp. CCQ). Suppose,  $\nu(\mathbb{P})$  is such that  $\nu(\mathbb{P}) = QS_{\mathcal{C}} = \langle \emptyset, \tau_{br}^{-1}(\chi(\mathbb{P})) \rangle$ . Intuitively,  $\mathcal{C}$  contains a context identifier  $c$ , for each predicate symbol  $c \in P$ . Also suppose,  $\nu_{cq}(Q) = \tau_{ccq}^{-1}(\chi(Q))$ . Notice that  $\nu_{cq}(Q)$  is CCQ. It can straightforwardly seen that  $\nu$  and  $\nu_{cq}$  can be computed in polynomial time, and  $\mathbb{P} \models_{fol} Q$  iff  $\nu(\mathbb{P}) \models \nu_{cq}(Q)$ .

Thanks to Theorem 12 and Theorem 14, the following theorem immediately holds:

**Theorem 15.** The CCQ EP over quad-systems is polynomially equivalent to CQ EP over ternary  $\forall\exists$  rule sets.

By virtue of the theorem above, we derive the following property:

*Property 6.* For quad-systems, the Quad EP, Quad-graph EP, BR(s) EP, and Quad-system EP are polynomially reducible to CCQ EP.

*Proof.* The following claim is a folklore in the realm of  $\forall\exists$  rules.

*Claim.* (1) The  $\forall\exists$  rule set EP is polynomially reducible to CQ EP.

Reducibility of  $\forall\exists$  rule EP to CQ EP is a folklore in the realm of  $\forall\exists$  rules. For a formal proof, we refer the reader to Baget et al. [15], where it is shown that the  $\forall\exists$  rule EP is polynomially reducible to fact (a set of instances) EP, and fact EP are equivalent to CQ EP. Also, Cali et al [34] show that CQ containment problem, which is equivalent to  $\forall\exists$  rule EP, is reducible to CQ EP. Since a  $\forall\exists$  rule set is a set of  $\forall\exists$  rules, by using a series of oracle calls to a function that solves the  $\forall\exists$  rule EP, we can define a function for deciding  $\forall\exists$  rule set entailment. Hence, the claim holds.

(a) Thanks to translation functions  $\tau$ ,  $\tau_{br}$  defined earlier, such that for any quad-system  $QS_C$ , quad-graph  $Q'_{C'}$ ,  $QS_C \models Q'_{C'}$  iff  $\tau(QS_C) \models_{fol} \tau_{br}(r_{Q'_{C'}})$ , we can infer that quad-graph EP is polynomially reducible to  $\forall\exists$  rule set EP. Applying claim 1, it follows the quad-graph EP over quad-systems is polynomially reducible to CQ EP over  $\forall\exists$  rule sets. By Theorem 14, we can deduce that quad-graph EP is polynomially reducible to CCQ EP.

(b) By the translation functions  $\tau$  and  $\tau_{br}$ , defined earlier, such that for any quad-system  $QS_C$ , a set of BRs  $R$ ,  $QS_C \models R$  iff  $\tau(QS_C) \models_{fol} \tau_{br}(R)$ , we can infer that BRs EP is polynomially reducible to  $\forall\exists$  rule set EP. Similar to (a) above, we deduce that BRs EP is polynomially reducible to CCQ EP.

From (a) and (b), it follows that Quad-system EP is reducible to CCQ EP.

Having seen that the CCQ EP over quad-systems is polynomially equivalent to CQ EP over ternary  $\forall\exists$  rule sets, we now compare some of the well known techniques used to ensure decidability of CQ entailment in the  $\forall\exists$  rules settings to the decidability techniques for quad-systems that we saw earlier in the previous sections. Note that since all the quad-system classes we proposed in this paper are FECs, for a judicious comparison, the  $\forall\exists$  rule classes to which we compare are classes which have a finite chase property. We compare to the following three well known classes: (i) Weakly Acyclic rule sets (WA), (ii) Jointly Acyclic rule sets (JA), and (iii) Model Faithful Acyclic  $\forall\exists$  rule sets (MFA). The following property is well known in the realm of  $\forall\exists$  rules:

*Property 7.* For the any  $\forall\exists$  rule set  $\mathbb{P}$ , the following holds:

1. If  $\mathbb{P} \in \text{WA}$ , then  $\mathbb{P} \in \text{JA}$  (from [38]),
2. If  $\mathbb{P} \in \text{JA}$ , then  $\mathbb{P} \in \text{MFA}$  (from [32]),
3.  $\text{WA} \subset \text{JA} \subset \text{MFA}$  (from [38] and [32]).

Note that a description of few other  $\forall\exists$  rule classes that do not have the finite chase property, but still enjoy decidability of CQ entailment are given in the related work.

## 6.1 Weak Acyclicity

Weak acyclicity [24, 26] is a popular technique used to detect whether a  $\forall\exists$  rule set has a finite chase, thus ensuring decidability of query answering. The set WA represents class of ternary  $\forall\exists$  rule sets that have the weak acyclicity property.

For any predicate atom  $p(t_1, \dots, t_n)$ , an expression  $\langle p, i \rangle$ , for  $i = 1, \dots, n$  is called a position of  $p$ . In the above case,  $t_1$  is said to occur at position  $\langle p, 1 \rangle$ ,  $t_2$  at  $\langle p, 2 \rangle$ , and so on. For a set of  $\forall\exists$  rules  $\mathbb{P}$ , its dependency graph is a graph whose nodes are positions of predicate atoms in  $\mathbb{P}$ ; for each  $r \in \mathbb{P}$  of the form (8), and for any variable  $x$  occurring in position  $\langle p, i \rangle$  in head of  $r$ :

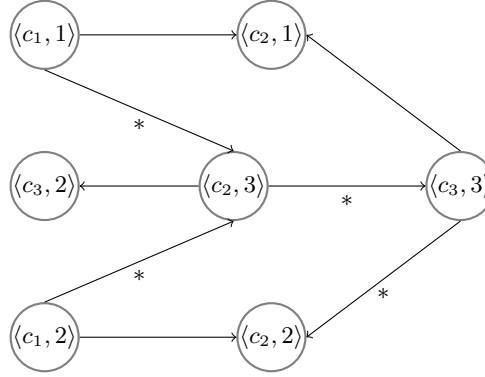


Fig. 3: Dependency graph of the quad-system in Example 1.

1. if  $x$  is universally quantified and  $x$  occurs in the body of  $r$  at position  $\langle p', j \rangle$ , then there exists an edge from  $\langle p', j \rangle$  to  $\langle p, i \rangle$
2. if  $x$  is existentially quantified, then for any universally quantified variable  $x'$  occurring in the head of  $r$ , with  $x'$  also occurring in the body of  $r$  at position  $\langle p', j \rangle$ , there exists a special edge from  $\langle p', j \rangle$  to  $\langle p, i \rangle$ .

$\mathbb{P}$  is called weakly acyclic, iff its dependency graph does not contain cycles going through a special edge. For any  $\forall\exists$  rule set  $\mathbb{P}$ , if  $\mathbb{P}$  is WA, then its chase is finite, and hence CQ EP is decidable. Note that the nodes in the dependency graph that has incoming special edges corresponds to the positions of predicates where new values are created due to existential variables, and the normal edges capture the propagation of constants from one predicate position to another predicate position. In this way, absence of cycles involving special edges ensures that newly created Skolem blank nodes are not recursively used to create other new Skolem blank nodes in the same position, leading to termination of chase computation.

*Example 3.* Let us revisit the quad-system  $QS_C = \langle Q_C, R \rangle$  mentioned in example 1, whose dependency graph is shown in Fig. 3. Note that the  $QS_C$  is unsafe, since its dChase contains a Skolem blank-node  $\_ : b_4$ , which has as descendant another Skolem blank node  $\_ : b_1$ , with the same origin context  $c_2$  (see Fig. 1). However, it can be seen from Fig. 3 that the dependency graph of  $\tau(QS_C)$  does not contain any directed cycle involving special edges. Hence  $\tau(QS_C)$  is weakly acyclic.

It turns out that there exists no inclusion relationship between the classes WA and CSAFE in either directions, i.e.  $WA \not\subseteq CSAFE$  (from example 3), and  $CSAFE \not\subseteq WA$  (from the fact that  $WA \subset JA$ , and example 4 below). Whereas  $WA \subset MSAFE$ , since  $WA \subset MFA$  and  $MFA \equiv MSAFE$  (Theorem 16).

## 6.2 Joint Acyclicity

Joint acyclicity [38] extends weak acyclicity, by also taking into consideration the join between variables in body of  $\forall\exists$  rules while analyzing the rules for acyclicity. The set

JA represents the class of all ternary  $\forall\exists$  rule sets that have the joint acyclicity property. A  $\forall\exists$  rule set  $\mathbb{P}$  is said to be *renamed apart*, if for any  $r \neq r' \in R$ ,  $\mathbf{V}(r) \cap \mathbf{V}(r') = \emptyset$ . Since any set of rules can be converted to an equivalent renamed apart one by simple variable renaming, we assume that any rule set  $\mathbb{P}$  is renamed apart. Also for any  $r \in \mathbb{P}$  and for a variable  $y$ , let  $Pos_H^r(y)$  ( $Pos_B^r(y)$ ) be the set of positions in which  $y$  occurs in the head (resp. body) of  $r$ . For any  $\forall\exists$  rule set  $\mathbb{P}$  and an existentially quantified variable  $y$  occurring in a rule in  $\mathbb{P}$ , we define  $Mov_{\mathbb{P}}(y)$  as the least set with:

- $Pos_H^r(y) \subseteq Mov_{\mathbb{P}}(y)$ , if  $y$  occurs in  $r$ ;
- $Pos_H^r(x) \subseteq Mov_{\mathbb{P}}(y)$ , if  $x$  is a universally quantified variable and  $Pos_B^r(x) \subseteq Mov_{\mathbb{P}}(y)$ ;

for any  $r \in \mathbb{P}$ . The existential dependency graph of a (renamed apart) set of rules  $\mathbb{P}$  is a graph whose nodes are the existentially quantified variables in  $\mathbb{P}$ . There exists an edge from a variable  $y$  to  $y'$ , if there is a rule  $r \in \mathbb{P}$  in which  $y'$  occurs and there exists a universally quantified variable  $x$  in the head (and body) of  $r$  such that  $Pos_B^r(x) \subseteq Mov_{\mathbb{P}}(y)$ . A  $\forall\exists$  rule set  $\mathbb{P}$  is jointly acyclic, iff its existential dependency graph is acyclic. Analyzing the containment relationships, it happens to be the case that  $JA \not\subseteq CSAFE$  (since  $WA \subset JA$ , and eg. 3). Also example 4 shows us that  $CSAFE \not\subseteq JA$ . However  $JA \subset MSAFE$ , since  $JA \subset MFA$  and  $MFA \equiv MSAFE$  (Theorem 16).

*Example 4.* Consider the quad-system  $QS_C = \langle Q_C, R \rangle$ , where  $Q_C = \{c_1: (a, b, c)\}$ . Suppose  $R$  is the following set:

$$R = \left\{ \begin{array}{l} c_1: (x_{11}, x_{12}, z_1) \rightarrow c_2: (x_{11}, x_{12}, y_1) \quad (r_1) \\ c_1: (x_{21}, x_{22}, z_2), c_2: (x_{22}, x_{21}, x_{23}) \rightarrow c_3: (x_{21}, x_{22}, x_{23}) \quad (r_2) \\ c_3: (x_{31}, x_{32}, x_{33}) \rightarrow c_1: (x_{33}, x_{31}, x_{32}) \quad (r_3) \end{array} \right\}$$

Iterations during the dChase construction are:

$$\begin{aligned} dChase_0(QS_C) &= \{c_1: (a, b, c)\} \\ dChase_1(QS_C) &= \{c_1: (a, b, c), c_2: (a, b, \_ : b_1)\} \\ dChase(QS_C) &= dChase_1(QS_C) \end{aligned}$$

Note that the lone Skolem blank node generated is  $\_ : b_1$ , which do not have any descendants. Hence, by definition  $QS_C$  is csafe (msafe/safe). Now analyzing the BRs for joint acyclicity, we note that for the only existentially quantified variable  $y_1$ ,

$$Mov_R(y_1) = \{\langle c_2, 3 \rangle, \langle c_3, 3 \rangle, \langle c_1, 1 \rangle\}$$

Since the BR  $r_1$  in which  $y_1$  occurs contains the universally quantified variable  $x_{11}$  in the head of  $r_1$  such that  $Pos_B^{r_1}(x_{11}) \subseteq Mov_R(y_1)$ , there exists a cycle from  $y_1$  to  $y_1$  itself in the existential dependency graph of  $\tau(QS_C)$ . Hence, by definition  $\tau(QS_C)$  is not joint acyclic. Also since the class of weakly acyclic rules are contained in the class of jointly acyclic rule, it follows that  $\tau(QS_C)$  is also not weakly acyclic.

### 6.3 Model Faithful Acyclicity (MFA)

MFA, proposed in Cuenca Grau et al. [32], is an acyclicity technique that guarantees finiteness of chase and decidability of query answering, in the realm of  $\forall\exists$  rules. The

set MFA denotes the class of all ternary  $\forall\exists$  rule sets that are model faithfully acyclic. As far as we know, the MFA technique subsumes almost all other known techniques that guarantee a finite chase, in the  $\forall\exists$  rules setting. Obviously,  $\text{WA} \subset \text{JA} \subset \text{MFA}$ .

For any  $\forall\exists$  rule  $r = \phi(r)(\mathbf{x}, \mathbf{z}) \rightarrow \psi(r)(\mathbf{x}, \mathbf{y})$ , for each  $y_j \in \{\mathbf{y}\}$ , let  $Y_r^j$  be a fresh unary predicate unique for  $y_j$  and  $r$ ; furthermore, let  $S$  be a fresh binary predicate. The transformation  $\mathbf{mfa}$  of  $r$  is defined as:

$$\mathbf{mfa}(r) = \phi(r)(\mathbf{x}, \mathbf{z}) \rightarrow \psi(r)(\mathbf{x}, \mathbf{y}) \wedge \bigwedge_{y_j \in \{\mathbf{y}\}} [Y_r^j(y_j) \wedge \bigwedge_{x_k \in \{\mathbf{x}\}} S(x_k, y_j)]$$

Also let  $r_1$  and  $r_2$  be two additional rules defined as:

$$\begin{aligned} S(x_1, z) \wedge S(z, x_2) &\rightarrow S(x_1, x_2) & (r_1) \\ Y_r^j(x_1) \wedge S(x_1, x_2) \wedge Y_r^j(x_2) &\rightarrow \mathfrak{C} & (r_2) \end{aligned}$$

where  $\mathfrak{C}$  is a fresh nullary predicate. For any set of  $\forall\exists$  rules  $\mathbb{P}$ , let  $ad(\mathbb{P})$  be the union of  $r_1$  with the set of rules obtained by instantiating  $r_2$ , for each  $r \in \mathbb{P}$ , for each existential variable  $y_j$  in  $r$ . For a set of  $\forall\exists$  rules  $\mathbb{P}$ ,  $\mathbf{mfa}(\mathbb{P}) = \bigcup_{r \in \mathbb{P}} \mathbf{mfa}(r) \cup ad(\mathbb{P})$ . A  $\forall\exists$  rule set  $\mathbb{P}$  is said to be MFA, iff  $\mathbf{mfa}(\mathbb{P}) \not\models_{\text{fol}} \mathfrak{C}$ . It was shown in Cuenca Grau et al. [32] that if  $\mathbb{P}$  is MFA, then  $\mathbb{P}$  has a finite chase, thus ensuring decidability of query answering. The following theorem establishes the fact that the notion of msafety is equivalent to MFA, thanks to the polynomial time translations between quad-systems and ternary  $\forall\exists$  rule sets.

**Theorem 16.** *Let  $\tau$  be the translation function from the set of unrestricted quad-systems to the set of ternary  $\forall\exists$  rule sets, as defined in Definition 8, then, for any quad-system  $QS_C = \langle Q_C, R \rangle$ ,  $QS_C$  is msafe iff  $\tau(QS_C)$  is MFA.*

*Proof.* (outline) Recall that  $\tau = \langle \tau_q, \tau_{br} \rangle$ , where  $\tau_q$  is the quad translation function and  $\tau_{br}$  is the translation function from BRs to  $\forall\exists$  rules. Also,  $\tau(QS_C) = \tau_{br}(\{r_{Q_C}\} \cup R)$ . Also, recall that for every blank node  $b$  in  $Q_C$ , the BR  $r_{Q_C}$  contains a corresponding existentially quantified variable  $y_b$ . We already saw that for such a transformation, the following property holds: for any  $m \in \mathbb{N}$ ,  $\tau_q(dChase_m(QS_C)) = chase_m(\tau(QS_C))$ , and for any BR  $r \in R \cup \{r_{Q_C}\}$ , an assignment  $\mu$ ,  $applicable_{R \cup \{r_{Q_C}\}}(r, \mu, dChase_m(QS_C))$  iff  $applicable_{\tau(QS_C)}(\tau_{br}(r), \mu, chase_m(\tau(QS_C)))$ . Also notice that for any two blank nodes  $\_ : b_1, \_ : b_2$ ,  $S(\_ : b_1, \_ : b_2) \in chase(\tau(QS_C))$ , iff  $\_ : b_1$  is a descendant of  $\_ : b_2$  in  $dChase(QS_C)$ . Hence, the relations  $S$  and descendantOf are identical.

Intuitively, MFA looks for cyclic creation of a Skolem blank-node whose descendant is another Skolem blank-node that is generated by the same rule  $r = body(r)(\mathbf{x}, \mathbf{z}) \rightarrow head(r)(\mathbf{x}, \mathbf{y})$ , by the same existential variable in  $y_j \in \{\mathbf{y}\}$  of  $r$ . Whereas, msafety looks only for generation of a Skolem blank-node  $\_ : b'$  whose descendant is another Skolem  $\_ : b$  using the same rule  $r$ . Hence, if  $\tau(QS_C)$  is not MFA, then  $QS_C$  is not msafe, and consequently onlyIf part of the theorem trivially holds.

(If part) Suppose  $QS_C$  is unmsafe, and  $\mu$  and  $\mu'$  are the assignments applied on  $r \in R$  to create Skolem blank nodes  $\_ : b$  and  $\_ : b'$ , respectively, and suppose  $\_ : b$  is a descendant of  $\_ : b'$  in the  $dChase(QS_C)$ . That is  $\_ : b = \mu(y_j)$  and  $\_ : b' = \mu'(y_k)$ , for  $y_j, y_k \in \{\mathbf{y}\}$  of  $r$ . Suppose  $j = k$ , then the prerequisite of non-MFA is trivially

satisfied. Suppose if  $j \neq k$  is the case, then there exists  $\_ : b''$  in  $dChase(QS_C)$  such that  $\_ : b'' = \mu'(y_j)$ , since  $\mu'$  is applied on  $r$  and  $y_j \in \{\mathbf{y}\}$ . This means that also in this case, the prerequisite of non-MFA is satisfied. As a consequence  $\tau(QS_C)$  is not MFA. Hence it follows that,  $QS_C$  is msafe iff  $\tau(QS_C)$  is MFA.

Let us revisit the quad-system  $QS_C$  in example 2, it can be easily seen that  $\tau(QS_C)$  is not MFA. Recall that we have seen that  $QS_C$  is safe but not msafe. We consider the Theorem 16 to be of importance, as it not only establishes the equivalence of MFA and msafety, but thanks to it and the translation  $\tau$ , it can be deduced that the technique of safety, which we presented earlier, (strictly) extends the MFA technique. As far as we know, the MFA class of  $\forall\exists$  rule sets is one of the most expressive class in the realm of  $\forall\exists$  rule sets which allows a finite chase. Hence, the notion of safety that we propose can straightforwardly be ported to  $\forall\exists$  settings. The main difference between MFA and safety is that MFA only looks for cyclic creation of two distinct Skolem blank-nodes  $\_ : b, \_ : b'$  that are generated by the same rule  $r$ , by the same existential variable in  $r$ . Whereas safety also takes into account the origin vectors  $\mathbf{a}$  and  $\mathbf{a}'$  used during rule application to create  $\_ : b$  and  $\_ : b'$ , respectively, and only raises an alarm if  $\mathbf{a} \cong \mathbf{a}'$ . Although, equivalence holds only between quad-systems and ternary  $\forall\exists$  rule sets, it can easily be noticed that the technique of safety can be applied to  $\forall\exists$  rule sets of arbitrary arity, and can be used to extend currently established tools and systems that work on existing notions of acyclicity such as WA, JA, or MFA.

## 7 Related Work

*Contexts and Distributed Logics* Work on contexts gained its attention as early as in the 80s, as McCarthy [1] proposed context as a solution to the generality problem in AI. After this, various studies about logics of contexts mainly in the field of KR were done by Guha [18], *Distributed First Order Logics* by Ghidini et al. [17] and *Local Model Semantics* by Giunchiglia et al. [8]. Primarily in these works, contexts were formalized as a first order/propositional theory and bridge rules were provided to inter-operate the various theories of contexts. Some of the initial works on contexts relevant to semantic web were the ones like *Distributed Description Logics* [5] by Borgida et al., and *Context-OWL* [7] by Bouquet et al., and the work of CKR [13, 10] by Serafini et al. These were mainly logics based on DLs, which formalized contexts as OWL KBs, whose semantics is given using a distributed interpretation structure with additional semantic conditions that suits varying requirements. Compared to these works, the bridge rules we consider are much more expressive with conjunctions and existential variables that supports value/blank-node creation.

*Temporal RDF/Annotated RDF* Studies in extending standard RDF with dimensions such as time and annotations have already been accomplished. Gutierrez et al. in [41] tried to add a temporal extension to RDF and defines the notion of a ‘temporal rdf graph’, in which a triple is augmented to a quadruple of the form  $t : (s, p, o)$ , where  $t$  is a time point. Whereas annotated extensions to RDF and querying annotated graphs have been studied in Udrea et al. [42] and Straccia et al. [43]. Unlike the case of time,



here the quadruple has the form:  $a : (s, p, o)$ , where  $a$  is an annotation. The authors provide semantics, inference rules and query language that allows for expressing temporal/annotated queries. Although these approaches, in a way address contexts by means of time and annotations, the main difference in our work is that we provide the means to specify expressive bridge rules for inter-operating the reasoning between the various contexts.

*DL+rules* Works on extending DL KBs with Datalog like rules was studied by Horrocks et al. [29] giving rise to the SWRL [29] language. Related initiatives propose a formalism using which one can mix a DL ontology with the Unary/Binary Datalog RuleML sublanguages of the Rule Markup Language, and hence enables Horn-like rules to be combined with an OWL KB. Since SWRL is undecidable in general, studies on computable sub-fragments gave rise to works like Description Logic Rules [40], where the authors deal with rules that can be totally internalized by a DL knowledge base, and hence if the DL considered is decidable, then also is a DL+rules KB. The authors give various fragments of the rule bases like SROIQ rules, EL++ rules etc. and show that certain new constructs that are not expressible by plain DL can be expressed using rules, although they are finally internalized into DL KBs. Unlike in our scenario, these works consider only horn rules without existential variables.

*$\forall\exists$  rules, TGDs, Datalog+- rules* Query answering over rules with universal-existential quantifiers in the context of databases, where these rules are called Datalog+- rules/tuple generating dependencies (TGDs), was done by Beeri and Vardi [14] even in the early 80s, where the authors show that the query entailment problem, in general, is undecidable. However, recently many classes of such rules have been identified for which query answering is decidable. These classes (according to [15]) can broadly be divided into the following three categories: (i) *bounded treewidth sets* (BTS), (ii) *finite unification sets* (FUS), and (iii) *finite extension sets* (FES). BTS contains the classes of  $\forall\exists$  rule sets, whose models have bounded treewidth. Some of the important classes of these sets are the linear  $\forall\exists$  rules [20], (weakly) guarded rules [34], (weakly) frontier guarded rules [15], and jointly frontier guarded rules [38]. BTS classes in general need not have a finite chase, and query answering is done by exploiting the fact that the chase is tree shaped, whose nodes (which are sets of instances) start replicating (up to isomorphism) after a while. Hence, one could stop the computation of the chase, once it can be made sure that any future iterations of chase can only produce nodes that are isomorphic to existing nodes. A deterministic algorithm for deciding query entailment for the greedy BTS, which is a subset of this class is provided in Thomazo et al. [16].

FUS classes include the class of ‘sticky’ rules [36, 35], atomic hypothesis rules in which the body of each rule contains only a single atom, and also the class of linear  $\forall\exists$  rules. The approach used for query answering in FUS classes is to rewrite the input query w.r.t. to the  $\forall\exists$  rule sets to another query that can be evaluated directly on the set of instances, such that the answers for the former query and latter query coincides. The approach is called the *query rewriting approach*. Compared to approaches proposed in this paper, these approaches do not enjoy the finite chase property, and are hence not conducive to materialization/forward chaining based query answering.

Unlike BTS and FUS, the FES classes are characterized by the finite chase property, and hence are most related to the techniques proposed in our work. Some of the classes in this set employ termination guarantying checks called ‘acyclicity tests’ that analyze the information flow between rules to check whether cyclic dependencies exist that can lead to infinite chase. *Weak acyclicity* [24, 26], was one of the first such notions, and was extended to *joint acyclicity* [38] and *super weak acyclicity* [37]. The main approach used in these techniques is to exploit the structure of the rules and use a dependency graph that models the propagation path of constants across various predicates in the rules, and restricting the dependency graph to be acyclic. The main drawback of these approaches is that they only analyze the schema/Tbox part of the rule sets, and ignore the instance part, and hence produce a large number of false alarms, i.e. it is often the case that although dependency graph is cyclic, the chase is finite. Recently, a more dynamic approach, called the MFA technique, that also takes into account the instance part of the rule sets was proposed in Cuenca Grau et al. [32], where existence of cyclic Skolem blank-node/constant generations in the chase is detected by augmenting the rules with extra information that keeps track of the Skolem function used to generate each Skolem blank-node. As shown in section 6, our technique of safety subsumes the MFA technique, and supports for much more expressive rule sets, by also keeping track of the vectors used by rule bodies while Skolem blank-nodes are generated.

*Data integration* Studies in query answering on integrated heterogeneous databases with expressive integration rules in the realm of data integration is primarily studied in the following two settings: (i) Data exchange [24], in which there is a source database and target database that are connected with existential rules, and (ii) Peer-to-peer data management systems (PDMS) [19], where there are an arbitrary number of peers that are interconnected using existential rules.

The approach based on dependency graphs, for instance, is used by Halevi et al. in the context of peer-peer data management systems [19], and decidability is attained by not allowing any kind of cycles in the peer topology. Whereas in the context of Data exchange, WA is used in [24, 26] to assure decidability, and the recent work by Marnette [37] employs the super weak acyclicity (SWA) to ensure decidability. It was shown in Cuenca Grau et al [32] that their MFA technique strictly subsumes both WA and SWA techniques in expressivity. Since we saw in section 6 that our technique of safety subsumes the MFA technique and allows the representation of much more expressive rule sets, the safety technique can straightforwardly be employed in the above mentioned systems with decidability guarantees for query answering.

## 8 Summary and Conclusion

In this paper, we study the problem of query answering over contextualized RDF knowledge in the presence of forall-existential bridge rules. We show that the problem, in general, is undecidable, and present a few decidable classes of quad-systems. Table 1 displays the complexity results of chase computation and query entailment for the various classes of quad-systems we have derived. Classes *csafe*, *msafe*, and *safe*, ensure decidability by restricting the structure of Skolem blank-nodes generated in the

Quad-System Fragment	Chase size w.r.t input quad-system	Data Complexity of CCQ entailment	Combined Complexity of CCQ entailment	Complexity of Recognition
Unrestricted Quad-Systems	Infinite	Undecidable	Undecidable	PTIME
Safe Quad-Systems	Double exponential	PTIME-complete	2EXPTIME-complete	2EXPTIME
MSafe Quad-Systems	Double exponential	PTIME-complete	2EXPTIME-complete	2EXPTIME
CSafe Quad-Systems	Double exponential	PTIME-complete	2EXPTIME-complete	2EXPTIME
RR Quad-Systems	Polynomial	PTIME-complete	EXPTIME	PTIME
Restricted RR Quad-Systems	Polynomial	PTIME-complete	NP-complete	PTIME

Table 1: Complexity info for various quad-system fragments

dChase. Briefly, the above classes do not allow an infinite descendant chain for Skolem blank-nodes generated, by constraining each Skolem blank-node in a descendant chain to have a different value for certain attributes, whose value sets are finite. RR and restricted RR quad-systems, do not allow the generation of Skolem blank nodes, thus constraining the dChase to have only constants from the initial quad-system. The above classes which suit varying situations, can be used to extend the currently established tools for contextual reasoning to give support for expressive bridge rules with conjunctions and existential quantifiers with decidability guarantees. From an expressivity point of view, the class of safe quad-systems subsumes all the above classes, and other well known classes in the realm of  $\forall\exists$  rules with finite chases. We view the results obtained in this paper as a general foundation for contextual reasoning and query answering over contextualized RDF knowledge formats such as quads, and can straightforwardly be used to extend existing quad stores.

## 9 Acknowledgements

We sincerely thank Loris Bozzatto (FBK-IRST, Italy), and Francesco Corcoglionitti (FBK-IRST, Italy), and Prof. Roberto Zunino (DISI, University of Trento, Italy) for all their helpful technical feedbacks on an initial version of this paper. We also thank Dr. Christoph Lange (School of Computer Science, University of Birmingham, UK), Prof. Sethumadhavan (Center for Cyber Security, Amrita University, India), and Prof. Padmanabhan T.R. (Dept. of Computer Science, Amrita University, India), for their time and motivating discussions.

## References

1. J. McCarthy, “Generality in AI,” *Communications of the ACM*, vol. 30, no. 12, pp. 1029–1035, 1987.
2. J. Carroll, C. Bizer, P. Hayes, and P. Stickler, “Named graphs, provenance and trust,” in Allan Ellis, Tatsuya Hagino (Eds.) *WWW ’05: Proceedings of the 14th international conference on World Wide Web*, (New York, NY, USA), pp. 613–622, ACM, 2005.
3. B. Schueler, S. Sizov, S. Staab, and D. T. Tran, “Querying for meta knowledge,” in Jinpeng Huai, Robin Chen, Hsiao-Wuen Hon, Yunhao Liu, Wei-Ying Ma, Andrew Tomkins, Xiaodong

- Zhang (Eds.) *WWW '08: Proceedings of the 17th international conference on World Wide Web*, (New York, NY, USA), pp. 625–634, ACM, 2008.
4. D.Lenat, “The Dimensions of Context Space,” tech. rep., CYCorp, 1998. Published online <http://www.cyc.com/doc/context-space.pdf>.
  5. A. Borgida and L. Serafini, “Distributed Description Logics: Assimilating Information from Peer Sources,” *Journal on Data Semantics*, vol. 1, Lecture Notes in Computer Science, Volume 2800, pp. 153–184, 2003.
  6. J. McCarthy, “Notes on Formalizing Context,” in *Proceedings of the 13th international joint conference on Artificial intelligence (ICAI' 95)* - Volume 1, Morgan Kaufmann Publishers Inc. Pages 555-560, 1995.
  7. P. Bouquet, F. Giunchiglia, F. van Harmelen, L. Serafini, and H. Stuckenschmidt. C-OWL: Contextualizing Ontologies. In Dieter Fensel, Katia P. Sycara, John Mylopoulos (Eds.): *The Semantic Web - ISWC 2003*, Springer 2003 Lecture Notes in Computer Science Volume, 2870, pages 164–179, 2003.
  8. F. Giunchiglia and C. Ghidini. Local models semantics, or contextual reasoning = locality + compatibility. *Artificial Intelligence*, Vol. 127, 2001, Pages 221–259.
  9. L. Bozzato, C. Ghidini and L. Serafini. Comparing contextual and flat representations of knowledge: a concrete case about football data. In V. Richard Benjamins, Mathieu d’Aquin, Andrew Gordon (Eds.) *Proceedings of the seventh international conference on Knowledge capture (K-CAP '13)*, 9-16, ACM 2013.
  10. M.Joseph and L.Serafini. Simple reasoning for contextualized RDF knowledge. In Oliver Kutz, Thomas Schneider (Eds.): *Modular Ontologies - Proceedings of the Fifth International Workshop (WOMO-2011)*, Volume 230, IOS Press, Frontiers in Artificial Intelligence and Applications, 2011.
  11. M.Joseph and G.Kuper and L.Serafini. Query Answering over contextualized RDF Knowledge with Forall-Existential Bridge Rules: Attaining Decidability using Acyclicity. In Laura Giordano, Valentina Gliozzi, Gian Luca Pozzato (Eds.): *Proceedings of the 29th Italian Conference in Computational Logic (CILC-2014)*, Volume 1195 of CEUR Workshop Proceedings, pages 210-224, CEUR-WS.org, 2014.
  12. M.Joseph and G.Kuper and L.Serafini. Query Answering over contextualized RDF/OWL Knowledge with Forall-Existential Bridge Rules: Attaining Decidability using Acyclicity. In Roman Kontchakov, Marie-Laure Mugnier (Eds.): *Web Reasoning and Rule Systems - 8th International Conference (RR-2014)*, Springer Lecture Notes in Computer Science Volume 8741, pp 60-75, 2014.
  13. L. Serafini and M. Homola. Contextualized knowledge repositories for the semantic web. *Web Semantics: Science, Services and Agents on the World Wide Web*, Special Issue: Reasoning with Context in the Semantic Web, Elsevier, Vol 12 - 13 (2012), 2012.
  14. C. Beeri and M. Y. Vardi. The Implication Problem for Data Dependencies. In Shimon Even, Oded Kariv (Eds.): *Automata, Languages and Programming*, 8th Colloquium, *ICALP*, Springer Lecture Notes in Computer Science Volume 115, pp 73-85, 1981.
  15. Baget, J.-F.; Leclère, M.; Mugnier, M.-L.; and Salvat, E. 2011. On rules with existential variables: Walking the decidability line. *Artificial Intelligence* 175(9-10):1620–1654.
  16. Thomazo, M.; Baget, J.-F.; Mugnier, M.-L.; and Rudolph, S. A Generic Querying Algorithm for Greedy Sets of Existential Rules. In Gerhard Brewka, Thomas Eiter, Sheila A. McIlraith (Eds.): *KR'12: International Conference on Principles of Knowledge Representation and Reasoning*, 096–106. 2012.
  17. C. Ghidini and L. Serafini. Distributed first order logics. In *Frontiers Of Combining Systems 2, Studies in Logic and Computation*, pages 121–140. Research Studies Press, 1998.
  18. R.Guha. *Contexts: a Formalization and some Applications*. PhD thesis, Stanford, 1992.

19. A. Y. Halevy, Z. G. Ives, D. Suciu, and I. Tatarinov, "Schema mediation in peer data management systems," in Umeshwar Dayal, Krithi Ramamritham, T. M. Vijayaraman (Eds.): 19th International Conference on Data Engineering *In ICDE*, IEEE Computer Society, pp. 505–516, 2003.
20. D. S. Johnson and A. C. Klug, "Testing containment of conjunctive queries under functional and inclusion dependencies," *Journal of Computer and System Sciences*, vol. 28, no. 1, pp. 167–189, 1984.
21. S. Abiteboul, R. Hull, and V. Vianu, *Foundations of Databases*. Addison-Wesley, 1995.
22. O. Kutz, C. Lutz, F. Wolter, and M. Zakharyashev, "E-Connections of Abstract Description Systems," *Artificial Intelligence*, vol. 156, no. 1, pp. 1–73, 2004.
23. S. Klarman and V. Gutiérrez-Basulto, "Two-dimensional description logics for context-based semantic interoperability," in Wolfram Burgard, Dan Roth (Eds.): *Proceedings of Twenty-Fifth AAAI Conference on Artificial Intelligence (AAAI-11)*, AAAI Press 2011.
24. R. Fagin, P. G. Kolaitis, R. J. Miller, and L. Popa, "Data Exchange: Semantics and Query Answering," in *Theoretical Computer Science*, pp. 28(1):89–124, 2005.
25. D. Calvanese, G. Giacomo, D. Lembo, M. Lenzerini, R. Rosati, "Tractable Reasoning and Efficient Query Answering in Description Logics: The DL-Lite Family", In *Journal of Automated Reasoning*, Springer-Verlag New York, Inc., pp. 385–429, 2007.
26. A. Deutsch and V. Tannen, "Reformulation of XML Queries and Constraints," in Diego Calvanese, Maurizio Lenzerini, Rajeev Motwani (Eds.): *In Proceedings of the 9th International Conference on Database Theory (ICDT' 03)*, Springer Lecture Notes in Computer Science, pp. 225–241, 2003.
27. H. J. ter Horst, "Completeness, decidability and complexity of entailment for RDF Schema and a semantic extension involving the OWL vocabulary," *Web Semantics: Science, Services and Agents on the WWW*, vol. 3, no. 2-3, pp. 79–115, 2005. Selected Papers from the ISWC, 2004.
28. B. Glimm, C. Lutz, I. Horrocks, and U. Sattler, "Answering conjunctive queries in the *SHIQ* description logic," in Manuela M. Veloso (Ed.) *Proceedings of the Twentieth International Joint Conference on Artificial Intelligence (IJCAI-07)*, pp. 299–404, AAAI Press, 2007.
29. I. Horrocks, P. F. Patel-Schneider, H. Boley, S. Tabet, B. Grosz, and M. Dean, "SWRL: A Semantic Web Rule Language Combining OWL and RuleML," W3C member submission, World Wide Web Consortium, 2004. <http://www.w3.org/Submission/SWRL/>
30. A. Deutsch, A. Nash, and J. Remmel, "The chase revisited," in Maurizio Lenzerini, Domenico Lembo (Eds.): *Proceedings of the twenty-seventh ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*, PODS '08, (New York, NY, USA), pp. 149–158, ACM, 2008. -
31. P. Hayes, ed., *RDF Semantics*. W3C Recommendation, Feb. 2004. <http://www.w3.org/TR/2004/REC-rdf-mt-20040210/>
32. B. Cuenca Grau, I. Horrocks, M. Krötzsch, C. Kupke, D. Magka, B. Motik, and Z. Wang, "Acyclicity Notions for Existential Rules and Their Application to Query Answering in Ontologies," in *Journal of Artificial Intelligence Research (JAIR)*, vol. 47, pp. 741–808, AI Access Foundation, 2013.
33. M. A. Harrison, *Introduction to Formal Language Theory*. Boston, MA, USA: Addison-Wesley Longman Publishing Company, Inc., 1st ed., 1978.
34. A. Cali, G. Gottlob, and M. Kifer Taming the infinite chase: Query answering under expressive relational constraints. In Gerhard Brewka, Jérôme Lang (Eds.): *Principles of Knowledge Representation and Reasoning (KR' 08)*, 70–80. 2008.
35. A. Cali, G. Gottlob, and A. Pieris, "Towards more expressive ontology languages: The query answering problem," in *Artificial Intelligence*, vol. 93, pp. 87–128, Elsevier, 2012.

36. A. Cali, G. Gottlob, and A. Pieris, “Query Answering under Non-guarded Rules in Datalog+/-,” in Pascal Hitzler, Thomas Lukasiewicz (Eds.) *Web Reasoning and Rule Systems - Fourth International Conference (RR’ 10)*, vol. 6333 of Springer Lecture Notes in Computer Science, pp. 1–17, Springer, 2010.
37. B. Marnette, “Generalized schema-mappings: from termination to tractability,” in Jan Paredaens, Jianwen Su (Eds.): *Proceedings of the twenty-eighth ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*, PODS ’09, (New York, NY, USA), pp. 13–22, ACM, 2009.
38. M. Krötzsch and S. Rudolph, “Extending decidable existential rules by joining acyclicity and guardedness,” in Toby Walsh (Ed.) *Proceedings of the 22nd International Joint Conference on Artificial Intelligence (IJCAI’11)* (T. Walsh, ed.), pp. 963–968, AAAI Press/IJCAI, 2011.
39. M. Krötzsch and S. Rudolph, “Conjunctive Queries for EL with Role Composition” in Diego Calvanese, Enrico Franconi, Volker Haarslev, Domenico Lembo, Boris Motik, Anni-Yasmin Turhan, Sergio Tessaris (Eds.): *Proceedings of the 20th International Workshop on Description Logics (DL’07)* (D. Calvanese, ed.), Vol. 250, CEUR Workshop Proceedings, 2007.
40. M. Krötzsch, S. Rudolph, and Pascal Hitzler. Description logic rules. In Malik Ghallab, Constantine D. Spyropoulos, Nikos Fakotakis, Nikolaos M. Avouris (Eds.): *Proceedings of the 18th European Conference on Artificial Intelligence (ECAI’08)*, pages 80–84. IOS Press, Frontiers in Artificial Intelligence and Applications, 2008.
41. C. Gutierrez, C. A. Hurtado, and A. A. Vaisman, “Temporal rdf,” in Lora Aroyo, Grigoris Antoniou, Eero Hyvnen, Annette ten Teije, Heiner Stuckenschmidt, Liliana Cabral, Tania Tudorache (Eds.): *2nd Annual European Semantic Web Conference (ESWC’ 05)*, pp. 93–107, Springer, Lecture Notes in Computer Science, 2005.
42. O. Udrea, D. R. Recupero, and V. S. Subrahmanian, “Annotated RDF,” *ACM Transactions in Computational Logic*, vol. 11, no. 2, pp. 1–41, 2010.
43. U. Straccia, N. Lopes, G. Lukacsy, and A. Polleres, “A general framework for representing and reasoning with annotated semantic web data,” in Maria Fox, David Poole (Eds.): *Proceedings of the 24th AAAI Conference on Artificial Intelligence (AAAI 2010), Special Track on Artificial Intelligence and the Web*, July 2010.

## A Proofs for Section 3

*Proof (Property 1).* Note that a strict linear order is a relation that is irreflexive, transitive, and linear.

**Irreflexivity:** By contradiction, suppose  $\prec_q$  is not irreflexive, then there exists  $Q \in \mathcal{Q}$  such that  $Q \prec_q Q$  holds. This means that neither of the conditions (i) and (ii) of  $\prec_q$  definition holds for  $Q$ . Hence, due to condition (iii)  $Q \not\prec_q Q$ , which is a contradiction.

**Linearity:** Note that for any two distinct  $Q, Q' \in \mathcal{Q}$ , one of the following holds: (a)  $Q \subset Q'$ , (b)  $Q' \subset Q$ , or (c)  $Q \setminus Q'$  and  $Q' \setminus Q$  are non-empty and disjoint. Suppose (a) is the case, then  $Q \prec_q Q'$  holds. Similarly, if (b) is the case then  $Q' \prec_q Q$  holds. Otherwise if (c) is the case, then by condition (ii), either  $Q \prec_q Q'$  or  $Q' \prec_q Q$  should hold. Hence,  $\prec_q$  is a linear order over  $\mathcal{Q}$ .

**Transitivity:** Suppose there exists  $Q, Q', Q'' \in \mathcal{Q}$  such that  $Q \prec_q Q'$  and  $Q' \prec_q Q''$ . Then, one of the following four cases hold: (a)  $Q \prec_q Q'$  due to (i) and  $Q' \prec_q Q''$  due to (i), (b)  $Q \prec_q Q'$  due to (i) and  $Q' \prec_q Q''$  due to (ii), (c)  $Q \prec_q Q'$  due to (ii) and  $Q' \prec_q Q''$  due to (i), (d)  $Q \prec_q Q'$  due to (ii) and  $Q' \prec_q Q''$  due to (ii).

Suppose if (a) is the case, then trivially  $Q \subset Q''$ , and hence by applying condition (i)  $Q \prec_q Q''$ . Otherwise if (b) is the case, then either (1)  $Q \subset Q''$  or (2)  $Q \not\subset Q''$ .

Suppose, (1) is the case then, by (i)  $Q \prec_q Q''$ . Otherwise, if (2) is the case, then since,  $Q \subset Q'$ , it cannot be the case that  $\text{greatestQuad}_{\prec_l}(Q'' \setminus Q) \prec_l \text{greatestQuad}_{\prec_l}(Q'' \setminus Q')$ , and it cannot be the case that  $\text{greatestQuad}_{\prec_l}(Q' \setminus Q'') \prec_l \text{greatestQuad}_{\prec_l}(Q \setminus Q'')$ . Hence, it should be the case that  $\text{greatestQuad}_{\prec_l}(Q'' \setminus Q') \preceq_l \text{greatestQuad}_{\prec_l}(Q'' \setminus Q)$  and  $\text{greatestQuad}_{\prec_l}(Q \setminus Q'') \prec_l \text{greatestQuad}_{\prec_l}(Q' \setminus Q'')$ . But since,  $\text{greatestQuad}_{\prec_l}(Q' \setminus Q'') \prec_l \text{greatestQuad}_{\prec_l}(Q'' \setminus Q')$ , it allows us to derive  $\text{greatestQuad}_{\prec_l}(Q \setminus Q'') \prec_l \text{greatestQuad}_{\prec_l}(Q'' \setminus Q)$ , and hence by condition (ii),  $Q \prec_q Q''$ . Hence, if (b) is the case, then in both possible cases (1) or (2), it should be the case that  $Q \prec_q Q''$ . Otherwise if (c) is the case, then similar to the arguments in (b), by condition (i) or (ii), it can easily be seen that  $Q \prec_q Q''$ .

Otherwise, if (d) is the case, then the following must hold:  $\text{greatestQuad}_{\prec_l}(Q \setminus Q') \prec_l \text{greatestQuad}_{\prec_l}(Q' \setminus Q)$  ( $\dagger$ ) and  $\text{greatestQuad}_{\prec_l}(Q' \setminus Q'') \prec_l \text{greatestQuad}_{\prec_l}(Q'' \setminus Q')$  ( $\ddagger$ ). Suppose by contradiction  $Q'' \prec_q Q$ , then one of the following holds: (1)  $Q'' \prec_q Q$  by condition (i) or (2)  $Q'' \prec_q Q$  by condition (ii). Suppose, (1) is the case, then it should be the case that  $Q'' \subset Q$ . Hence, it should not be the case that  $\text{greatestQuad}_{\prec_l}(Q \setminus Q') \prec_l \text{greatestQuad}_{\prec_l}(Q'' \setminus Q')$  and it should not be the case that  $\text{greatestQuad}_{\prec_l}(Q' \setminus Q'') \prec_l \text{greatestQuad}_{\prec_l}(Q' \setminus Q)$ . Hence, it should be the case that  $\text{greatestQuad}_{\prec_l}(Q'' \setminus Q') \preceq_l \text{greatestQuad}_{\prec_l}(Q \setminus Q')$  ( $\heartsuit$ ), and it should be the case that  $\text{greatestQuad}_{\prec_l}(Q' \setminus Q) \preceq_l \text{greatestQuad}_{\prec_l}(Q' \setminus Q'')$  ( $\spadesuit$ ). Applying ( $\ddagger$ ) in ( $\heartsuit$ ), we get  $\text{greatestQuad}_{\prec_l}(Q' \setminus Q'') \prec_l \text{greatestQuad}_{\prec_l}(Q \setminus Q')$ , and Applying ( $\dagger$ ) in ( $\spadesuit$ ), we get  $\text{greatestQuad}_{\prec_l}(Q \setminus Q') \prec_l \text{greatestQuad}_{\prec_l}(Q' \setminus Q'')$ , which is a contradiction. Suppose if (2) is the case, then  $\text{greatestQuad}_{\prec_l}(Q'' \setminus Q) \prec_l \text{greatestQuad}_{\prec_l}(Q \setminus Q'')$ . The above can be written as:  $\text{greatestQuad}_{\prec_l}(Q'' \setminus (Q \cap Q'')) \prec_l \text{greatestQuad}_{\prec_l}(Q \setminus (Q \cap Q''))$ . Using  $Q \cap Q' \cap Q'' \subseteq Q \cap Q'$ , it follows that  $\text{greatestQuad}_{\prec_l}(Q'' \setminus (Q \cap Q' \cap Q'')) \preceq_l \text{greatestQuad}_{\prec_l}(Q \setminus (Q \cap Q' \cap Q''))$  ( $\clubsuit$ ). Also applying similar transformation in ( $\dagger$ ) and ( $\ddagger$ ), we get  $\text{greatestQuad}_{\prec_l}(Q \setminus (Q \cap Q' \cap Q'')) \preceq_l \text{greatestQuad}_{\prec_l}(Q' \setminus (Q \cap Q' \cap Q''))$ , and  $\text{greatestQuad}_{\prec_l}(Q' \setminus (Q \cap Q' \cap Q'')) \preceq_l \text{greatestQuad}_{\prec_l}(Q'' \setminus (Q \cap Q' \cap Q''))$ . From which, it follows that  $\text{greatestQuad}_{\prec_l}(Q \setminus (Q \cap Q' \cap Q'')) \preceq_l \text{greatestQuad}_{\prec_l}(Q'' \setminus (Q \cap Q' \cap Q''))$ . Using ( $\clubsuit$ ) in the above, we get  $\text{greatestQuad}_{\prec_l}(Q \setminus (Q \cap Q' \cap Q'')) = \text{greatestQuad}_{\prec_l}(Q' \setminus (Q \cap Q' \cap Q'')) = \text{greatestQuad}_{\prec_l}(Q'' \setminus (Q \cap Q' \cap Q''))$ , which is a contradiction. Hence, it should be the case that  $Q \prec_q Q''$ .

*Proof (Theorem 2).* We show that CCQ entailment is undecidable for unrestricted quad-systems, by showing that the well known undecidable problem of “non-emptiness of intersection of context-free grammars” is reducible to the CCQ answering problem.

Given an alphabet  $\Sigma$ , string  $w$  is a sequence of symbols from  $\Sigma$ . A language  $L$  is a subset of  $\Sigma^*$ , where  $\Sigma^*$  is the set of all strings that can be constructed from the alphabet  $\Sigma$ , and also includes the empty string  $\epsilon$ . Grammars are machineries that generate a particular language. A grammar  $G$  is a quadruple  $\langle V, T, S, P \rangle$ , where  $V$  is the set of variables,  $T$ , the set of terminals,  $S \in V$  is the start symbol, and  $P$  is a set of production rules (PR), in which each PR  $r \in P$ , is of the form:

$$w \rightarrow w'$$

where  $w, w' \in \{T \cup V\}^*$ . Intuitively application of a PR  $r$  of the form above on a string  $w_1$ , replaces every occurrence of the sequence  $w$  in  $w_1$  with  $w'$ . PRs are applied starting from the start symbol  $S$  until it results in a string  $w$ , with  $w \in \Sigma^*$  or no more

production rules can be applied on  $w$ . In the former case, we say that  $w \in L(G)$ , the language generated by grammar  $G$ . For a detailed review of grammars, we refer the reader to Harrison et al. [33]. A *context-free grammar* (CFG) is a grammar, whose set of PRs  $P$ , have the following property:

*Property 8.* For a CFG, every PR is of the form  $v \rightarrow w$ , where  $v \in V$ ,  $w \in \{T \cup V\}^*$ .

Given two CFGs,  $G_1 = \langle V_1, T, S_1, P_1 \rangle$  and  $G_2 = \langle V_2, T, S_2, P_2 \rangle$ , where  $V_1, V_2$  are the set of variables,  $T$  such that  $T \cap (V_1 \cup V_2) = \emptyset$  is the set of terminals.  $S_1 \in V_1$  is the start symbol of  $G_1$ , and  $P_1$  are the set of PRs of the form  $v \rightarrow w$ , where  $v \in V$ ,  $w$  is a sequence of the form  $w_1 \dots w_n$ , where  $w_i \in V_1 \cup T$ .  $S_2, P_2$  are defined similarly. Deciding whether the language generated by the grammars  $L(G_1)$  and  $L(G_2)$  have non-empty intersection is known to be undecidable [33].

Given two CFGs,  $G_1 = \langle V_1, T, S_1, P_1 \rangle$  and  $G_2 = \langle V_2, T, S_2, P_2 \rangle$ , we encode grammars  $G_1, G_2$  into a quad-system of the form  $QS_c = \langle Q_c, R \rangle$ , with a single context identifier  $c$ . Each PR  $r = v \rightarrow w \in P_1 \cup P_2$ , with  $w = w_1 w_2 w_3 \dots w_n$ , is encoded as a BR of the form:

$$c: (x_1, w_1, x_2), c: (x_2, w_2, x_3), \dots, c: (x_n, w_n, x_{n+1}) \rightarrow c: (x_1, v, x_{n+1}) \quad (10)$$

where  $x_1, \dots, x_{n+1}$  are variables. W.l.o.g. we assume that the set of terminal symbols  $T$  is equal to the set of terminal symbols occurring in  $P_1 \cup P_2$ . For each terminal symbol  $t_i \in T$ ,  $R$  contains a BR of the form:

$$c: (x, \text{rdf:type}, C) \rightarrow \exists y \ c: (x, t_i, y), c: (y, \text{rdf:type}, C) \quad (11)$$

and  $Q_c$  contains only the triple:

$$c: (a, \text{rdf:type}, C)$$

We in the following show that:

$$QS_c \models \exists y \ c: (a, S_1, y) \wedge c: (a, S_2, y) \leftrightarrow L(G_1) \cap L(G_2) \neq \emptyset \quad (12)$$

*Claim.* (1) For any  $w = t_1, \dots, t_p \in T^*$ , there exists  $b_1, \dots, b_p$ , such that  $c: (a, t_1, b_1)$ ,  $c: (b_1, t_2, b_2), \dots, c: (b_{p-1}, t_p, b_p), c: (b_p, \text{rdf:type}, C) \in dChase(QS_c)$ .

we proceed by induction on  $|w|$ .

**base case** suppose if  $|w| = 1$ , then  $w = t_i$ , for some  $t_i \in T$ . But by construction  $c: (a, \text{rdf:type}, C) \in dChase_0(QS_c)$ , on which rules of the form (11) is applicable. Hence, there exists an  $i$  such that  $dChase_i(QS_c)$  contains  $c: (a, t_i, b_i)$ ,  $c: (b_i, \text{rdf:type}, C)$ , for each  $t_i \in T$ . Hence, the base case.

**hypothesis** for any  $w = t_1 \dots t_p$ , if  $|w| \leq p'$ , then there exists  $b_1, \dots, b_p$ , such that  $c: (a, t_1, b_1), c: (b_1, t_2, b_2), \dots, c: (b_{p-1}, t_p, b_p), c: (b_p, \text{rdf:type}, C) \in dChase(QS_c)$ .

**inductive step** suppose  $w = t_1 \dots t_{p+1}$ , with  $|w| \leq p' + 1$ . Since  $w$  can be written as  $w' t_{p+1}$ , where  $w' = t_1 \dots t_p$ , and by hypothesis, there exists  $b_1, \dots, b_p$  such that  $c: (a, t_1, b_1), c: (b_1, t_2, b_2), \dots, c: (b_{p-1}, t_p, b_p), c: (b_p, \text{rdf:type}, C) \in dChase(QS_c)$ . Also since rules of the form (11) are applicable on  $c: (b_p, \text{rdf:type}, C)$ , and hence produces triples of the form  $c: (b_p, t_i, b_{p+1}^i), c: (b_{p+1}^i, \text{rdf:type}, C)$ , for each  $t_i \in T$ . Since  $t_{p+1} \in T$ , the claim follows.



For a grammar  $G = \langle V, T, S, P \rangle$ , whose start symbol is  $S$ , and for any  $w \in \{V \cup T\}^*$ , for some  $V_j \in V$ , we denote by  $V_j \rightarrow^i w$ , the fact that  $w$  was derived from  $V_j$  by  $i$  production steps, i.e. there exists steps  $V_j \rightarrow r_1, \dots, r_i \rightarrow w$ , which lead to the production of  $w$ . For any  $w, w \in L(G)$ , iff there exists an  $i$  such that  $S \rightarrow^i w$ . For any  $V_j \in V$ , we use  $V_j \rightarrow^* w$  to denote the fact that there exists an arbitrary  $i$ , such that  $V_j \rightarrow^i w$ .

*Claim.* (2) For any  $w = t_1 \dots t_p \in \{V \cup T\}^*$ , and for any  $V_j \in V$ , if  $V_j \rightarrow^* w$  and there exists  $b_1, \dots, b_{p+1}$ , with  $c: (b_1, t_1, b_2), \dots, c: (b_p, t_p, b_{p+1}) \in dChase(QS_c)$ , then  $c: (b_1, V_j, b_{p+1}) \in dChase(QS_c)$ .

We prove this by induction on the size of  $w$ .

**base case** Suppose  $|w| = 1$ , then  $w = t_k$ , for some  $t_k \in T$ . If there exists  $b_1, b_2$  such that  $c: (b_1, t_k, b_2)$ . But since there exists a PR  $V_j \rightarrow t_k$ , by transformation given in (10), there exists a BR  $c: (x_1, t_k, x_2) \rightarrow c: (x_1, V_j, x_2) \in R$ , which is applicable on  $c: (b_1, t_k, b_2)$  and hence the quad  $c: (b_1, V_j, b_2) \in dChase(QS_c)$ .

**hypothesis** For any  $w = t_1 \dots t_p$ , with  $|w| \leq p'$ , and for any  $V_j \in V$ , if  $V_j \rightarrow^* w$  and there exists  $b_1, \dots, b_{p+1}$ , such that  $c: (b_1, t_1, b_2), \dots, c: (b_p, t_p, b_{p+1}) \in dChase(QS_c)$ , then  $c: (b_1, V_j, b_{p+1}) \in dChase(QS_c)$ .

**inductive step** Suppose if  $w = t_1 \dots t_{p+1}$ , with  $|w| \leq p' + 1$ , and  $V_j \rightarrow^i w$ , and there exists  $b_1, \dots, b_{p+1}, b_{p+2}$ , such that  $c: (b_1, t_1, b_2), \dots, c: (b_{p+1}, t_{p+1}, b_{p+2}) \in dChase(QS_c)$ . Also, one of the following holds (i)  $i = 1$ , or (ii)  $i > 1$ . Suppose (i) is the case, then it is trivially the case that  $c: (b_1, V_j, b_{p+2}) \in dChase(QS_c)$ . Suppose if (ii) is the case, one of the two sub cases holds (a)  $V_j \rightarrow^{i-1} V_k$ , for some  $V_k \in V$  and  $V_k \rightarrow^1 w$  or (b) there exist a  $V_k \in V$ , such that  $V_k \rightarrow^* t_{q+1} \dots t_{q+l}$ , with  $2 \leq l \leq p$ , where  $V_j \rightarrow^* t_1 \dots t_q V_k t_{p-l+1} \dots t_{p+1}$ . If (a) is the case, trivially then  $c: (b_1, V_k, b_{q+2}) \in dChase(QS_c)$ , and since by construction there exists  $c: (x_0, V_k, x_1) \rightarrow c: (x_0, V_{k+1}, x_1), \dots, c: (x_0, V_{k+i}, x_1) \rightarrow c: (x_0, V_j, x_1) \in R$ ,  $c: (b_1, V_j, b_{q+2}) \in dChase(QS_c)$ . If (b) is the case, then since  $|t_{q+1} \dots t_{q+l}| \geq 2$ ,  $|t_1 \dots t_q V_k t_{p-l+1} \dots t_{p+1}| \leq p'$ . This implies that  $c: (b_1, V_j, b_{p+2}) \in dChase(QS_c)$ .

Similarly, by construction of  $dChase(QS_c)$ , the following claim can straightforwardly be shown to hold:

*Claim.* (3) For any  $w = t_1 \dots t_p \in \{V \cup T\}^*$ , and for any  $V_j \in V$ , if there exists  $b_1, \dots, b_p, b_{p+1}$ , with  $c: (b_1, t_1, b_2), \dots, c: (b_p, t_p, b_{p+1}) \in dChase(QS_c)$  and  $c: (b_1, V_j, b_{p+1}) \in dChase(QS_c)$ , then  $V_j \rightarrow^* w$ .

(a) For any  $w = t_1 \dots t_p \in T^*$ , if  $w \in L(G_1) \cap L(G_2)$ , then by claim 1, since there exists  $b_1, \dots, b_p$ , such that  $c: (a, t_1, b_1), \dots, c: (b_{p-1}, t_p, b_p) \in dChase(QS_c)$ . But since  $w \in L(G_1)$  and  $w \in L(G_2)$ ,  $S_1 \rightarrow w$  and  $S_2 \rightarrow w$ . Hence by claim 2,  $c: (a, S_1, b_p), c: (a, S_2, b_p) \in dChase(QS_c)$ , which implies that  $dChase(QS_c) \models \exists y c: (a, s_1, y) \wedge c: (a, s_2, y)$ . Hence, by Theorem 1,  $QS_c \models \exists y c: (a, s_1, y) \wedge c: (a, s_2, y)$ . (b) Suppose if  $QS_c \models \exists y c: (a, S_1, y) \wedge c: (a, S_2, y)$ , then applying Theorem 1, it follows that there exists  $b_p$  such that  $c: (a, S_1, b_p), c: (a, S_2, b_p) \in dChase(QS_c)$ . Then it is the case that there exists  $w = t_1 \dots t_p \in T^*$ , and  $b_1, \dots, b_p$  such that  $c: (a, t_1, b_1), \dots, c: (b_{p-1}, t_p, b_p), c: (a, S_1, b_p), c: (a, S_2, b_p) \in dChase(QS_c)$ . Then by claim 3,  $S_1 \rightarrow^* w, S_2 \rightarrow^* w$ . Hence,  $w \in L(G_1) \cap L(G_2)$ .

By (a),(b) it follows that there exists  $w \in L(G_1) \cap L(G_2)$  iff  $QS_C \models \exists y c: (a, s_1, y) \wedge c: (a, s_2, y)$ . As we have shown that the intersection of CFGs, which is an undecidable problem, is reducible to the problem of query entailment on unrestricted quad-system, the latter is undecidable.

## B Proofs for Section 4

*Proof (Theorem 5).* We in the following show the case of  $dChase^{csafe}(QS_C)$ , i.e. **unC-Safe**  $\in dChase^{csafe}(QS_C)$  iff  $QS_C$  is unsafe. The proof follows from Lemma 6 and Lemma 7 below.

The proofs for the case of  $dChase^{safe}(QS_C)$  and  $dChase^{msafe}(QS_C)$  is similar, and is omitted.

**Lemma 6 (Soundness).** *For any quad-system  $QS_C = \langle Q_C, R \rangle$ , if the quad **unC-Safe**  $\in dChase^{csafe}(QS_C)$ , then  $QS_C$  is unsafe.*

*Proof.* Note that  $augC(R) = \bigcup_{r \in R} augC(r) \cup \{brTR\}$ , where  $brTR$  is the range restricted BR  $c_c: (x_1, \text{descendantOf}, z), c_c: (z, \text{descendantOf}, x_2) \rightarrow c_c: (x_1, \text{descendantOf}, x_2)$ . Also for each  $r \in R$ ,  $body(r) = body(augC(r))$ , and for any  $c \in \mathcal{C}$ ,  $c: (s, p, o) \in head(r)$  iff  $c: (s, p, o) \in head(augC(r))$ . That is,  $head(r) = head(augC(r))(\mathcal{C})$ , where  $head(r)(\mathcal{C})$  denotes the quad-patterns in  $head(r)$ , whose context identifiers is in  $\mathcal{C}$ . Also,  $head(augC(r)) = head(augC(r))(\mathcal{C}) \cup head(augC(r))(c_c)$ , and also the set of existentially quantified variables in  $head(augC(r))(c_c)$  is contained in the set of existentially quantified variables in  $head(augC(r))(\mathcal{C})$  ( $\dagger$ ). We first prove the following claim:

*Claim.* (0) For any quad-system  $QS_C = \langle Q_C, R \rangle$ , let  $i$  be a csafe dChase iteration, let  $j$  be the number of csafe dChase iterations before  $i$  in which  $brTR$  was applied, then  $dChase_{i-j}(QS_C) = dChase_i^{csafe}(QS_C)(\mathcal{C})$ .

We approach the proof of the above claim by induction on  $i$ .

**base case** If  $i = 1$ , then  $dChase_0^{csafe}(QS_C)(c_c) = \emptyset$  and  $dChase_0^{csafe}(QS_C)(\mathcal{C}) = dChase_0^{csafe}(QS_C) = dChase_0(QS_C)$ . Hence, it should be the case that  $applicable_{augC(R)}(brTR, \mu, dChase_0^{csafe}(QS_C))$  does not hold, for any  $\mu$ . Hence,  $applicable_R(r, \mu, dChase_0(QS_C))$  iff  $applicable_{augC(R)}(augC(r), \mu, dChase_0^{csafe}(QS_C))$ , for any  $r \in R$ , assignment  $\mu$ . Also using ( $\dagger$ ), it follows that  $dChase_1(QS_C) = dChase_{1-0}^{csafe}(QS_C)(\mathcal{C})$ .

**hypothesis** for any  $i \leq k$ , if  $i$  is a csafe dChase iteration, and  $j$  be the number of csafe dChase iterations before  $i$  in which  $brTR$  was applied, then  $dChase_{i-j}(QS_C) = dChase_i^{csafe}(QS_C)(\mathcal{C})$ .

**inductive** suppose  $i = k + 1$ , then one of the following three cases should hold: (a)  $applicable_{augC(R)}(r, \mu, dChase_k^{csafe}(QS_C))$  does not hold for any  $r \in augC(R)$ , assignment  $\mu$ , and  $dChase_{k+1}^{csafe}(QS_C) = dChase_k^{csafe}(QS_C)$ , or (b)  $applicable_{augC(R)}(brTR, \mu, dChase_k^{csafe}(QS_C))$  holds, for some assignment  $\mu$ , or (c)  $applicable_{augC(R)}(r, \mu, dChase_k^{csafe}(QS_C))$  holds, for some  $r \in augC(R) \setminus \{brTR\}$ , for some assignment  $\mu$ . If (a) is the case, then it should be the case that  $applicable_R(r',$

$\mu, dChase_{k-j}(QS_C)$  does not hold, for any  $r' \in R$ , assignment  $\mu$ . As a result  $dChase_{k+1-j}(QS_C) = dChase_{k-j}(QS_C)$ , and hence,  $dChase_{k+1-j}(QS_C) = dChase_{k+1}^{csafe}(QS_C)(C)$ . If (b) is the case, then since  $dChase_{k+1}^{csafe}(QS_C)(C) = dChase_k^{csafe}(QS_C)(C)$ ,  $dChase_{k+1}^{csafe}(QS_C)(C) = dChase_{k+1-j-1}(QS_C) = dChase_{k-j}(QS_C)$ . If (c) is the case, then it should be the case that  $applicable_R(r', \mu, dChase_{k-j}(QS_C))$ , where  $r = augC(r')$  and  $head(r)(C) = head(r)$ . Hence, it should be the case that  $dChase_{k+1}^{csafe}(QS_C)(C) = dChase_{k+1-j}(QS_C)$ .

The following claim, which straightforwardly follows from claim 0, shows that any quad  $c: (s, p, o)$ , with  $c \in C$  derived in csafe dChase, is also derived in its standard dChase. In this way, csafe dChase do not generate any unsound triples in any context  $c \in C$ .

*Claim.* (1) For any quad  $c: (s, p, o)$ , where  $c \in C$ , if  $c: (s, p, o) \in dChase^{csafe}(QS_C)$ , then  $c: (s, p, o) \in dChase(QS_C)$ .

The following claim shows that the set of origin context quads are also sound.

*Claim.* (2) If there exists quad  $c_c: (b, originContext, c) \in dChase^{csafe}(QS_C)$ , then  $c \in originContexts(b)$ .

If  $c_c: (b, originContext, c) \in dChase^{csafe}(QS_C)$ , there exists  $i \in \mathbb{N}$ , such that  $c_c: (b, originContext, c) \in dChase_i^{csafe}(QS_C)$  and there exists no  $j < i$  with  $c_c: (b, originContext, c) \in dChase_j^{csafe}(QS_C)$ . But if  $c_c: (b, originContext, c) \in dChase_i^{csafe}(QS_C)$  implies that there exists an  $augC(r) = body(x, z) \rightarrow head(x, y) \in augC(R)$ , with  $c_c: (y_j, originContext, c) \in head(x, y)$ ,  $y_j \in \{y\}$ , such that  $c_c: (b, originContext, c)$  was generated due to application of an assignment  $\mu$  on  $augC(r)$ , with  $b = y_j[\mu^{ext}(y)]$ . This implies that there exists  $c: (s, p, o) \in head(x, y)$ , with  $s = y_j$  or  $p = y_j$  or  $o = y_j$ ,  $c \in C$ . Since according to our assumption,  $i$  is the first iteration in which  $c_c: (b, originContext, c)$  is generated, it follows that  $i$  is the first iteration in which  $c: (s, p, o)[\mu^{ext}(y)]$  is also generated. Let  $k$  be the number of iterations before  $i$  in which  $brTR$  was applied. By applying claim 0, it should be the case that  $c: (s, p, o)[\mu^{ext}(y)] \in dChase_{i-k}(QS_C)$ , and  $i - k$  should be the first such dChase iteration. Hence,  $c \in originContexts(b)$ .

In the following claim, we prove the soundness of the descendant quads generated in a safe dChase.

*Claim.* (3) For any two distinct blank nodes  $b, b'$  in  $dChase^{csafe}(QS_C)$ , if  $c_c: (b', descendantOf, b) \in dChase^{csafe}(QS_C)$  then  $b'$  is a descendant of  $b$ .

Since any quad of the form  $c_c: (b', descendantOf, b) \in dChase^{csafe}(QS_C)$  is not an element of  $Q_C$ , and can only be introduced by an application of a BR  $r \in augC(R)$ , any quad of the form  $c_c: (b', descendantOf, b)$  can only be introduced, earliest in the first iteration of  $dChase^{csafe}(QS_C)$ . Suppose  $c_c: (b', descendantOf, b) \in dChase_j^{csafe}(QS_C)$ , then there exists an iteration  $i \geq 1$  such that  $c_c: (b', descendantOf, b) \in dChase_j^{csafe}(QS_C)$ , for any  $j \geq i$ , and  $c_c: (b', descendantOf, b) \notin dChase_{j'}^{csafe}(QS_C)$ , for any  $j' < i$ . We apply induction on  $i$  for the proof.

**base case** suppose  $c_c: (b', \text{descendantOf}, b) \in dChase_{-e_1}^{\text{csafe}}(QS_C)$  and since  $b \neq b'$ , then there exists a BR  $r \in \text{augC}(R)$ ,  $\exists \mu$  such that  $\text{applicable}_{\text{augC}(R)}(r, \mu, dChase_0^{\text{csafe}}(QS_C))$ , i.e.  $\text{body}(r)(\mathbf{x}, \mathbf{z})[\mu] \subseteq dChase_0^{\text{csafe}}(QS_C)$  and  $c_c: (b', \text{descendantOf}, b) \in \text{head}(r)(\mathbf{x}, \mathbf{y})[\mu^{\text{ext}(\mathbf{y})}]$ . Then by construction of  $\text{augC}(r)$ , it follows that  $b = y_j[\mu^{\text{ext}(\mathbf{y})}]$ , for some  $y_j \in \{\mathbf{y}\}$  and  $b' = \mu(x_i)$ , for some  $x_i \in \{\mathbf{x}\}$ . Since  $dChase_0(QS_C) = dChase_0^{\text{csafe}}(QS_C)$ , it follows using  $(\dagger)$  that  $\text{applicable}_R(r', \mu, dChase_{-e_0}(QS_C))$  holds, for  $r' = \text{body}(r')(\mathbf{x}, \mathbf{z}) \rightarrow \text{head}(r')(\mathbf{x}, \mathbf{y})$ , with  $\text{augC}(r') = r$ . Hence, by construction, it follows that  $b = y_j[\mu^{\text{ext}(\mathbf{y})}] \in \mathbf{C}(dChase_1(QS_C))$ , for  $y_j \in \{\mathbf{y}\}$  and  $b' = \mu(x_i)$ , for  $x_i \in \{\mathbf{x}\}$ . Hence  $b'$  is a descendant of  $b$  (by definition).

**hypothesis** if  $c_c: (b', \text{descendantOf}, b) \in dChase_i^{\text{csafe}}(QS_C)$ , for  $1 \leq i \leq k$ , then  $b'$  is a descendant of  $b$ .

**inductive step** suppose  $c_c: (b', \text{descendantOf}, b) \in dChase_{k+1}^{\text{csafe}}(QS_C)$ , then either (i)  $c_c: (b', \text{descendantOf}, b) \in dChase_k^{\text{csafe}}(QS_C)$  or (ii)  $c_c: (b', \text{descendantOf}, b) \notin dChase_k^{\text{csafe}}(QS_C)$ . Suppose (i) is the case, then by hypothesis,  $b'$  is a descendant of  $b$ . If (ii) is the case, then either (a)  $c_c: (b', \text{descendantOf}, b)$  is the result of the application of a  $brTR \in \text{augC}(R)$  on  $dChase_k^{\text{csafe}}(QS_C)$  or (b)  $c_c: (b', \text{descendantOf}, b)$  is the result of the application of a  $r \in \text{augC}(R) \setminus \{brTR\}$  on  $dChase_k^{\text{csafe}}(QS_C)$ . If (a) is the case, then there exists a  $b'' \in \mathbf{C}(dChase_k^{\text{csafe}}(QS_C))$  such that  $c_c: (b', \text{descendantOf}, b'') \in dChase_k^{\text{csafe}}(QS_C)$  and  $c_c: (b'', \text{descendantOf}, b) \in dChase_k^{\text{csafe}}(QS_C)$ . Hence, by hypothesis  $b'$  is a descendantOf  $b''$  and  $b''$  is a descendantOf  $b$ . Since ‘descendantOf’ relation is transitive,  $b'$  is a descendantOf  $b$ . Otherwise if (b) is the case then similar to the arguments used in the base case, it can easily be seen that  $b'$  is a descendant of  $b$ .

Suppose if the quad **unCSafe**  $\in dChase^{\text{csafe}}(QS_C)$ , then this implies that there exists an iteration  $i$  such that the function **unCSafeTest** on  $\text{augC}(r)$ , with  $r = \text{body}(r)(\mathbf{x}, \mathbf{z}) \rightarrow \text{head}(r)(\mathbf{x}, \mathbf{y}) \in R$ , assignment  $\mu$ , and  $dChase_i^{\text{csafe}}(QS_C)$  returns True. This implies that, there exists  $b, b' \in \mathbf{B}$ ,  $y_j \in \{\mathbf{y}\}$  such that  $\text{body}(r)(\mathbf{x}, \mathbf{z})[\mu] \subseteq dChase_i^{\text{csafe}}(QS_C)$ ,  $b \in \{\mu(\mathbf{x})\}$ ,  $c_c: (b', \text{descendantOf}, b) \in dChase_i^{\text{csafe}}(QS_C)$  and  $\{c \mid c_c: (b', \text{originContext}, c) \in dChase_i^{\text{csafe}}(QS_C)\} = cScope(y_j, \text{head}(r)(\mathbf{x}, \mathbf{y}))$ . Suppose  $k$  be the number of csafe dChase iterations before  $i$ , in which  $brTR$  was applied. Hence, by claim 0,  $dChase_{i-k-1}(QS_C) = dChase_{i-1}^{\text{csafe}}(QS_C)(\mathcal{C})$ , and consequently  $\text{applicable}_R(r, \mu, dChase_{i-k-1}(QS_C))$  holds. Hence, as a result of  $\mu$  being applied on  $r$ , there exists  $b'' = y_j[\mu^{\text{ext}(\mathbf{y})}] \in \mathbf{B}(dChase_{i-k}(QS_C))$ , with  $b \in \{\mu(\mathbf{x})\}$ . Hence, by definition  $\text{originContext}(b'') = cScope(y_j, \text{head}(r))$ , and  $b$  is a descendantOf  $b''$ . If  $b \neq b'$ , then by Claim 2,  $b'$  is a descendantOf  $b$ , otherwise  $b' = b$  and hence  $b'$  is a descendantOf  $b''$ . Consequently,  $b'$  is a descendantOf  $b''$ . Also, applying claim 3, we get that  $\text{originContexts}(b') = \text{originContexts}(b'')$ , which means that prerequisites of unsafety is satisfied, and hence,  $QS_C$  is unsafe.

**Lemma 7 (Completeness).** For any quad-system,  $QS_C = \langle Q_C, R \rangle$ , if  $QS_C$  is unsafe then **unCSafe**  $\in dChase^{\text{csafe}}(QS_C)$ .

*Proof.* We first prove a few supporting claims in order to prove the theorem.

*Claim.* (0) For any quad-system  $QS_C = \langle Q_C, R \rangle$ , suppose  $\mathbf{unCSafe} \notin dChase^{csafe}(QS_C)$ , then for any dChase iteration  $i$ , there exists a  $j \geq 0$  such that  $dChase_i(QS_C) = dChase_{i+j}^{csafe}(QS_C)(C)$ .

We approach the proof by induction on  $i$ .

**base case** for  $i = 0$ , we know that  $dChase_0(QS_C) = dChase_0^{csafe}(QS_C) = Q_C$ .

Hence, the base case trivially holds.

**hypothesis** for  $i \leq k \in \mathbb{N}$ , there exists  $j \geq 0$  such that  $dChase_i(QS_C) = dChase_{i+j}^{csafe}(QS_C)$

**step case** for  $i = k+1$ , one of the following holds: (a)  $dChase_{k+1}(QS_C) = dChase_k(QS_C)$  or (b)  $dChase_{k+1}(QS_C) = dChase_k(QS_C) \cup head(r)(\mathbf{x}, \mathbf{y})[\mu^{ext}(\mathbf{y})]$  and  $applicable_R(r, \mu, dChase_k(QS_C))$  holds, for some  $r = body(r)(\mathbf{x}, \mathbf{z}) \rightarrow head(r)(\mathbf{x}, \mathbf{y})$ , assignment  $\mu$ . If (a) is the case, then trivially the claim holds. Otherwise, if (b) is the case, then let  $j \in \mathbb{N}$  be such that  $dChase_k(QS_C) = dChase_{k+j}^{csafe}(QS_C)(C)$ . Let  $j' \geq j, l \in \mathbb{N}$  be such that  $applicable_{augC(R)}(brTR, \mu, dChase_{k+l}^{csafe}(QS_C))$ , for any  $j' \geq l \geq j$ , and  $applicable_{augC(R)}(brTR, \mu, dChase_{k+j'+1}^{csafe}(QS_C))$  does not hold. By construction, it should be the case that  $applicable(r', \mu, dChase_{k+j'+1}^{csafe}(QS_C))$  holds, where  $r' = augC(r)$ . Also since no new Skolem blank node was introduced in any csafe dChase iteration  $k+l$ , for any  $j \leq l \leq j'$ . It should be the case that  $head(r)[\mu^{ext}(\mathbf{y})] = head(r')[\mu^{ext}(\mathbf{y})](C)$ . Since,  $dChase_{k+l}^{csafe}(QS_C)(C) = dChase_k(QS_C)$ , for any  $j \leq l \leq j'$ , and  $dChase_{k+j'+1}^{csafe}(QS_C) = dChase_{k+j'}^{csafe}(QS_C) \cup head(r')[\mu^{ext}(\mathbf{y})]$ ,  $dChase_{k+j'+1}^{csafe}(QS_C)(C) = dChase_{k+1}(QS_C)$ . Hence, the claim follows.

The following claim, which straightforwardly follows from claim 0, shows that, for csafe quad-systems its standard dChase is contained in its safe dChase.

*Claim.* (1) Suppose  $\mathbf{unCSafe} \notin dChase^{csafe}(QS_C)$ , then  $dChase(QS_C) \subseteq dChase^{csafe}(QS_C)$ .

Claim below shows that the generation of originContext quads in csafe dChase is complete.

*Claim.* (2) For any quad-system  $QS_C$ , if  $\mathbf{unCSafe} \notin dChase^{csafe}(QS_C)$ , then for any Skolem blank-node  $b$  generated in  $dChase(QS_C)$ , and for any  $c \in C$ , if  $c \in originContexts(b)$ , then there exists a quad  $c_c: (b, originContext, c) \in dChase^{csafe}(QS_C)$ .

Since the only way a Skolem blank node  $b$  gets generated in any iteration  $i$  of  $dChase(QS_C)$  is by the application of a BR  $r \in R$ , i.e. when there  $\exists r = body(r)(\mathbf{x}, \mathbf{z}) \rightarrow head(r)(\mathbf{x}, \mathbf{y}) \in R$ , assignment  $\mu$ , such that  $applicable_R(r, \mu, dChase_{i-1}(QS_C))$ , and  $b = y_j[\mu^{ext}(\mathbf{y})]$ , for some  $y_j \in \{\mathbf{y}\}$ , and  $dChase_i(QS_C) = dChase_{i-1}(QS_C) \cup head(r)(\mathbf{x}, \mathbf{y})[\mu^{ext}(\mathbf{y})]$ . Also since  $c \in originContexts(b)$ , it should be the case that  $c \in cScope(y_j, head(r))$ . From claim 0, we know that there exists  $j \geq 0$ , such that  $dChase_i(QS_C) = dChase_{i+j}^{csafe}(QS_C)(C)$ . W.l.o.g, assume that  $i+j$  is the first such csafe dChase iteration. Hence, it follows that  $applicable_{augC(R)}(r', \mu, dChase_{i+j-1}^{csafe}(QS_C))$ , where  $r' = augC(r)$ . Since,  $head(r) \subseteq head(r')$ , it should be the case that  $c \in cScope(y_j, head(r'))$ . Hence, by construction of  $augC$ ,  $c_c: (y_j, originContext, c) \in$

$head(r')$ , and as a result of application of  $\mu$  on  $r'$  in iteration  $i+j$ ,  $c_c: (b, \text{originContext}, c)$  gets generated in  $dChase_{i+j}^{\text{csafe}}(QS_C)$ . Hence, the claim holds.

For the claim below, we introduce the concept of the sub-distance. For any two blank nodes, their sub-distance is inductively defined as:

**Definition 9.** For any two blank nodes  $b, b'$ ,  $\text{sub-distance}(b, b')$  is defined inductively as:

- $\text{sub-distance}(b, b') = 0$ , if  $b' = b$ ;
- $\text{sub-distance}(b, b') = \infty$ , if  $b \neq b'$  and  $b$  is not a descendant of  $b'$ ;
- $\text{sub-distance}(b, b') = \min_{t \in \{x[\mu]\}} \{ \text{sub-distance}(b, t) \} + 1$ , if  $b'$  was generated by application of  $\mu$  on  $r = \text{body}(r)(x, z) \rightarrow \text{head}(r)(x, y)$ , i.e.  $b' = y_j[\mu^{\text{ext}}(y)]$ , for some  $y_j \in \{y\}$ , and  $b$  is a descendant of  $b'$ .

*Claim.* (3) For any quad-system  $QS_C = \langle Q_C, R \rangle$ , if  $\text{unCSafe} \notin dChase^{\text{csafe}}(QS_C)$ , then for any two Skolem blank nodes  $b, b'$  in  $dChase(QS_C)$ , if  $b$  is a descendant of  $b'$  then there exists a quad of the form  $c_c: (b, \text{descendantOf}, b') \in dChase^{\text{csafe}}(QS_C)$ .

Note by the definition of sub-distance that if  $b$  is a descendant of  $b'$ , then  $\text{sub-distance}(b, b') \in \mathbb{N}$ . Assuming  $\text{unCSafe} \notin dChase^{\text{csafe}}(QS_C)$ , and  $b$  is a descendant of  $b'$ , we approach the proof by induction on  $\text{sub-distance}(b, b')$ .

**base case** Suppose  $\text{sub-distance}(b, b') = 1$ , then this implies that there exists  $r = \text{body}(x, z) \rightarrow \text{head}(r)(x, y)$ , assignment  $\mu$  such that  $b'$  was generated due to application of  $\mu$  on  $r$ , i.e.  $b' = y_j[\mu^{\text{ext}}(y)]$ , for some  $y_j \in \{y\}$ , and  $b \in \{x[\mu]\}$ . This implies that there exists a  $dChase$  iteration  $i$  such that  $\text{applicable}_R(r, \mu, dChase_i(QS_C))$  and  $dChase_{i+1}(QS_C) = dChase_i(QS_C) \cup \text{apply}(r, \mu)$ . Since  $\text{unCSafe} \notin dChase^{\text{csafe}}(QS_C)$ , using claim 0,  $\exists k \geq i$  such that  $dChase_i(QS_C) = dChase_k^{\text{csafe}}(QS_C)(C)$ . W.l.o.g., let  $k$  be the first such csafe  $dChase$  iteration. This means that  $\text{applicable}_{\text{augC}(R)}(r', \mu, dChase_k^{\text{csafe}}(QS_C))$ , where  $r' = \text{augC}(r)$ , and  $dChase_{k+1}^{\text{csafe}} = dChase_k^{\text{csafe}}(QS_C) \cup \text{head}(r')[\mu^{\text{ext}}(y)]$ , and  $b, b' \in \text{head}(r')[\mu^{\text{ext}}(y)]$ ,  $b \in \{x[\mu]\}$ ,  $b' = y_j[\mu^{\text{ext}}(y)]$ . By construction of  $\text{augC}()$ , since there exists a quad-pattern  $c_c: (x_l, \text{descendantOf}, y_j) \in \text{head}(r')$ , for any  $x_l \in \{x\}$ ,  $y_j \in \{y\}$ , it follows that  $c_c: (b, \text{descendantOf}, b') \in dChase_{k+1}^{\text{csafe}}(QS_C)$ .

**hypothesis** Suppose  $\text{sub-distance}(b, b') \leq k$ ,  $k \in \mathbb{N}$ , then  $c_c: (b, \text{descendantOf}, b') \in dChase^{\text{csafe}}(QS_C)$ .

**inductive step** Suppose  $\text{sub-distance}(b, b') = k + 1$ , then there exists a  $b'' \neq b$ , assignment  $\mu$ , and BR  $r = \text{body}(r)(x, z) \rightarrow \text{head}(r)(x, y) \in R$  such that  $b'$  was generated due to the application of  $\mu$  or  $r$  with  $b'' \in \{x[\mu]\}$ , i.e.  $b' = y_j[\mu^{\text{ext}}(y)]$ , for  $y_j \in \{y\}$ , and  $b$  is a descendant of  $b''$ . This implies that  $\text{sub-distance}(b'', b') = 1$ , and  $\text{sub-distance}(b, b'') = k$ , and hence by hypothesis  $c_c: (b, \text{descendantOf}, b'') \in dChase^{\text{csafe}}(QS_C)$ , and  $c_c: (b'', \text{descendantOf}, b') \in dChase^{\text{csafe}}(QS_C)$ . Hence, by construction of csafe  $dChase$ ,  $c_c: (b, \text{descendantOf}, b') \in dChase^{\text{csafe}}(QS_C)$ .

Suppose  $QS_C$  is unsafe, then by definition, there exists a blank nodes  $b, b'$  in  $\mathbf{B}_{sk}(dChase(QS_C))$ , such that  $b$  is descendant of  $b'$ , and  $\text{originContexts}(b)$  is equal to  $\text{originContexts}(b')$ . By contradiction, if  $\text{unCSafe} \notin dChase^{\text{csafe}}(QS_C)$ , then by claim 1,  $dChase(QS_C) \subseteq dChase^{\text{csafe}}(QS_C)$ . Since by claim 2, for any  $c \in \text{originContexts}(b)$ ,

there exists quads of the form  $c_c: (b, \text{originContext}, c) \in dChase^{\text{csafe}}(QS_C)$  and for every  $c' \in \text{originContexts}(b')$ , there exists  $c_c: (b', \text{originContext}, c') \in dChase^{\text{csafe}}(QS_C)$ . Since  $\text{originContexts}(b) = \text{originContexts}(b')$ , it follows that  $\{c \mid c_c: (b, \text{originContext}, c) \in dChase^{\text{csafe}}(QS_C)\} = \{c' \mid c_c: (b', \text{originContext}, c') \in dChase^{\text{csafe}}(QS_C)\}$ . Also by claim 3, since  $b$  is a descendant of  $b'$ , there exists a quad of the form  $c_c: (b, \text{descendantOf}, b') \in dChase^{\text{csafe}}(QS_C)$ . But, by construction of  $dChase^{\text{csafe}}(QS_C)$ , it should be the case that there exist a  $b'' \in \mathbf{B}_{sk}(dChase^{\text{csafe}}(QS_C))$ ,  $r = \text{body}(r)(x, z) \rightarrow \text{head}(r)(x, y) \in \text{augC}(R)$ , assignment  $\mu$  such that  $b'$  was generated due to the application of  $\mu$  on  $r$ , i.e.  $b' = y_j[\mu^{ext}(y)]$  with  $b'' \in \{x[\mu]\}$ , and  $c_c: (b, \text{descendantOf}, b'') \in dChase^{\text{csafe}}(QS_C)$ . But, since  $\{c \mid c_c: (b, \text{originContext}, c) \in dChase^{\text{csafe}}(QS_C)\} = cScope(y_j, \text{head}(r_i))$ , the method **unCSafeTest**( $r, \mu, dChase_i^{\text{csafe}}(QS_C)$ ) should return True, for some  $i \in \mathbb{N}$ . Hence, it should be the case that **unCSafe**  $\in dChase^{\text{csafe}}(QS_C)$ , which is a contradiction to our assumption. Hence **unCSafe**  $\in dChase^{\text{csafe}}(QS_C)$ , if  $dChase(QS_C)$  is unsafe.

*Proof (Property 4). (Only If)* By definition,  $R$  is universally safe (resp. msafe, resp csafe) iff  $\langle Q_C, R \rangle$  is safe (resp. msafe, resp. csafe), for any quad-graph  $Q_C$ . Hence,  $\langle Q_C^{crit}, R \rangle$  is safe (resp. msafe, resp. csafe).

(If part) We give the proof for the case of safe quad-systems. The proof for the msafe and csafe case can be obtained by slight modification. In order to show that if  $\langle Q_C^{crit}, R \rangle$  is safe, then  $R$  is universally safe, we prove the contrapositive. That is we show that if there exists  $Q_C$  such that  $\langle Q_C, R \rangle$  is unsafe, then  $QS_C^{crit} = \langle Q_C^{crit}, R \rangle$  is unsafe. Suppose, there exists such an unsafe quad-system  $QS_C = \langle Q_C, R \rangle$ , we show how to incrementally construct a homomorphism  $h$  from constants in  $dChase(QS_C)$  to the constants in  $dChase(QS_C^{crit})$  such that for any Skolem blank node  $\_ : b$  in  $dChase(QS_C)$ , there exists a homomorphism from descendence graph of  $\_ : b$  to the descendence graph of  $h(\_ : b)$  in  $dChase(QS_C^{crit})$ . Suppose  $h$  is initialized as: for any constant  $c \in \mathbf{C}(QS_C)$ ,  $h(c) = \_ : b^{crit}$ , if  $c \in \mathbf{C}(QS_C) \setminus \mathbf{C}(QS_C^{crit})$ ; and  $h(c) = c$  otherwise. It can be noted that for any BR  $r = \text{body}(r)(x, z) \rightarrow \text{head}(r)(x, y) \in R$ , if  $\text{body}(r)[\mu] \subseteq dChase_0(QS_C)$  then  $\text{body}(r)[\mu][h] \subseteq dChase_0(QS_C^{crit})$ . Now it follows that for any  $i \in \mathbb{N}$ ,  $\text{level}(\text{body}(r)[\mu]) = 0$  if  $\text{applicable}(r, \mu, dChase_i(QS_C))$ , then there exists  $j \leq i$  such that  $\text{applicable}(r, h \circ \mu, dChase_j(QS_C^{crit}))$ . Let  $h$  be extended so that for any  $i \in \mathbb{N}$ , for any Skolem blank node  $\_ : b$  introduced in  $dChase_{i+1}(QS_C)$  while applying  $\mu$  on  $r$ , for existential variable  $y \in \{y\}$ , let  $h(\_ : b)$  be the blank node introduced in  $dChase_{j+1}(QS_C^{crit})$ , for the existential variable  $y$  while applying  $h \circ \mu$  on  $r$ . Hence, it follows that, for any  $i \in \mathbb{N}$ ,  $\text{applicable}_R(r, \mu, dChase_i(QS_C))$  implies there exists  $j \leq i$  such that  $\text{applicable}(r, h \circ \mu, dChase_j(QS_C^{crit}))$ , for any  $r, \mu$ . Also note that, for any Skolem blank node  $\_ : b$  generated in  $dChase_i(QS_C)$ , it can be noted that  $\lambda_r(\_ : b) = \lambda_r(h(\_ : b))$  and  $\lambda_c(\_ : b) = \lambda_c(h(\_ : b))$  and  $\lambda_v(\_ : b)[h] = \lambda_v(h(\_ : b))$ . Hence, it follows that for any Skolem blank node  $\_ : b$  in  $dChase(QS_C)$ ,  $h$  is a homomorphism from descendence graph of  $\_ : b$  to the descendence graph of  $h(\_ : b)$  in  $dChase(QS_C^{crit})$ . Hence, if there exists two Skolem blank nodes  $\_ : b, \_ : b'$  in  $dChase(QS_C)$ , with  $\_ : b'$  a descendant of  $\_ : b$  and  $\text{originRuleId}(\_ : b) = \text{originRuleId}(\_ : b')$  and  $\text{originVector}(\_ : b) \cong \text{originVector}(\_ : b')$ , then it follows that there exists  $h(\_ : b), h(\_ : b')$  in  $dChase(QS_C^{crit})$ , with  $h(\_ : b')$  descendant of  $h(\_ : b)$  and  $\text{originRuleId}(h(\_ : b)) =$

$originRuleId(h(\_ : b'))$  and  $originVector(h(\_ : b)) \cong originVector(h(\_ : b'))$ . Hence, it follows from the definition that  $QS_{\mathcal{C}}^{critic}$  is unsafe.